IN THE UNITED STATES PAPTENT AND TRADEMARK OFFICE

APPLICATION FOR PATENT OF:

Dominic P. BEHAN, Karin LEHMANN-BRUINSMA, Derek T. CHALMERS, Ruoping CHEN, Huong T. DANG, Martin J. GORE, Chen W. LIAW, , I-Lin LIN, Kevin P. LOWITZ and Carol A. WHITE

FOR:

NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED HUMAN G PROTEIN-COUPLED RECEPTORS

ARENA PHARMACEUTICALS, INC. 6166 Nancy Ridge Drive San Diego, CA 92121 AREN-0054 PATENT

NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED HUMAN G PROTEIN-COUPLED RECEPTORS

This patent application is a continuation-in-part of, and claims priority from, U.S. Serial Number 09/170,496, filed with the United States Patent and Trademark Office on October 13, 1998. This application also claims the benefit of priority from the following provisional applications, all filed via U.S. Express Mail with the United States Patent and Trademark Office on the indicated dates: U.S. Provisional Number 60/110,060, filed November 27, 1998; U.S. Provisional Number 60/120,416, filed February 16, 1999; U.S. Provisional Number 60/121,852, filed February 26, 1999 claiming benefit of U.S.

- Provisional Number 60/109,213, filed November 20, 1998; U.S. Provisional Number 60/123,944, filed March 12, 1999; U.S. Provisional Number 60/123,945, filed March 12, 1999; U.S. Provisional Number 60/123,948, filed March 12, 1999; U.S. Provisional Number 60/123,951, filed March 12, 1999; U.S. Provisional Number 60/123,946, filed March 12, 1999; U.S. Provisional Number 60/123,949, filed March 12, 1999; U.S.
- Provisional Number 60/152,524, filed September 3, 1999, claiming benefit of U.S.

 Provisional Number 60/151,114, filed August 27, 1999 and U.S. Provisional Number 60/108,029, filed November 12, 1998; U.S. Provisional Number 60/136,436, filed May 28,

AREN-0054 - 2 - PATENT

1999; U.S. Provisional Number 60/136,439, filed May 28, 1999; U.S. Provisional Number 60/136,567, filed May 28, 1999; U.S. Provisional Number 60/137,127, filed May 28, 1999; U.S. Provisional Number 60/137,131, filed May 28, 1999; U.S. Provisional Number 60/141,448, filed June 29, 1999 claiming benefit of U.S. Provisional Number 60/136,437, filed May 28, 1999; U.S. Provisional Number 60/156,633, filed September 29, 1999; U.S. Provisional Number 60/156,555, filed September 29, 1999; U.S. Provisional Number 60/156,634, filed September 29, 1999; U.S. Provisional Number (Arena Pharmaceuticals, Inc. docket number: CHN10-1), filed September 29, 1999; U.S. Provisional Number ___ (Arena Pharmaceuticals, Inc. docket number: RUP6-1), filed October 1, 1999; U.S. Provisional Number ___(Arena Pharmaceuticals, Inc. docket number: RUP7-1), filed October 1, 1999; U.S. Provisional Number (Arena Pharmaceuticals, Inc. docket number: CHN6-1), filed October 1, 1999; U.S. Provisional Number __(Arena Pharmaceuticals, Inc. docket number: RUP5-1), filed October 1, 1999; and U.S. Provisional Number (Arena Pharmaceuticals, Inc. docket number: CHN9-1), filed October 1, 1999. This application is also related to co-pending U.S. Serial Number (Woodcock, Washburn, Kurtz, Makiewicz & Norris, LLP docket number AREN-0050), filed on October 12, 1999 (via U.S. Express Mail) and U.S. Serial Number 09/364,425, filed on July 30, 1999, both incorporated herein by reference. Each of the foregoing applications are incorporated by reference herein in their entirety.

FIELD OF THE INVENTION

The invention disclosed in this patent document relates to transmembrane receptors, and more particularly to human G protein-coupled receptors, and specifically to

GPCRs that have been altered to establish or enhance constitutive activity of the receptor. Preferably, the altered GPCRs are used for the direct identification of candidate compounds as receptor agonists, inverse agonists or partial agonists having potential applicability as therapeutic agents.

BACKGROUND OF THE INVENTION

5

15

Han men men

Although a number of receptor classes exist in humans, by far the most abundant and therapeutically relevant is represented by the G protein-coupled receptor (GPCR or GPCRs) class. It is estimated that there are some 100,000 genes within the human genome, and of these, approximately 2%, or 2,000 genes, are estimated to code for GPCRs. Receptors, including GPCRs, for which the endogenous ligand has been identified are referred to as "known" receptors, while receptors for which the endogenous ligand has not been identified are referred to as "orphan" receptors. GPCRs represent an important area for the development of pharmaceutical products: from approximately 20 of the 100 known GPCRs, 60% of all prescription pharmaceuticals have been developed.

GPCRs share a common structural motif. All these receptors have seven sequences of between 22 to 24 hydrophobic amino acids that form seven alpha helices, each of which spans the membrane (each span is identified by number, *i.e.*, transmembrane-1 (TM-1), transmebrane-2 (TM-2), etc.). The transmembrane helices are joined by strands of amino acids between transmembrane-2 and transmembrane-3, transmembrane-4 and transmembrane-5, and transmembrane-6 and transmembrane-7 on the exterior, or "extracellular" side, of the cell membrane (these are referred to as "extracellular" regions 1, 2 and 3 (EC-1, EC-2 and EC-3), respectively). The transmembrane helices are also joined by strands of amino acids between transmembrane-1 and transmembrane-2, transmembrane-3 and transmembrane-4, and

transmembrane-5 and transmembrane-6 on the interior, or "intracellular" side, of the cell membrane (these are referred to as "intracellular" regions 1, 2 and 3 (IC-1, IC-2 and IC-3), respectively). The "carboxy" ("C") terminus of the receptor lies in the intracellular space within the cell, and the "amino" ("N") terminus of the receptor lies in the extracellular space outside of the cell.

Generally, when an endogenous ligand binds with the receptor (often referred to as "activation" of the receptor), there is a change in the conformation of the intracellular region that allows for coupling between the intracellular region and an intracellular "G-protein." It has been reported that GPCRs are "promiscuous" with respect to G proteins, *i.e.*, that a GPCR can interact with more than one G protein. *See*, Kenakin, T., 43 *Life Sciences* 1095 (1988). Although other G proteins exist, currently, Gq, Gs, Gi, Gz and Go are G proteins that have been identified. Endogenous ligand-activated GPCR coupling with the G-protein begins a signaling cascade process (referred to as "signal transduction"). Under normal conditions, signal transduction ultimately results in cellular activation or cellular inhibition. It is thought that the IC-3 loop as well as the carboxy terminus of the receptor interact with the G protein.

Under physiological conditions, GPCRs exist in the cell membrane in equilibrium between two different conformations: an "inactive" state and an "active" state. A receptor in an inactive state is unable to link to the intracellular signaling transduction pathway to produce a biological response. Changing the receptor conformation to the active state allows linkage to the transduction pathway (via the G-protein) and produces a biological response.

A receptor may be stabilized in an active state by an endogenous ligand or a

15

AREN-0054 - 5 - PATENT

compound such as a drug. Recent discoveries, including but not exclusively limited to modifications to the amino acid sequence of the receptor, provide means other than endogenous ligands or drugs to promote and stabilize the receptor in the active state conformation. These means effectively stabilize the receptor in an active state by simulating the effect of an endogenous ligand binding to the receptor. Stabilization by such ligand-independent means is termed "constitutive receptor activation."

SUMMARY OF THE INVENTION

Disclosed herein are non-endogenous versions of endogenous, human GPCRs and uses thereof.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a representation of 8XCRE-Luc reporter plasmid (see, Example 4(c)3.)

Figures 2A and 2B are graphic representations of the results of ATP and ADP binding to endogenous TDAG8 (2A) and comparisons in serum and serum free media (2B).

Figure 3 is a graphic representation of the comparative signaling results of CMV versus the GPCR Fusion Protein H9(F236K):Gsa.

DETAILED DESCRIPTION

The scientific literature that has evolved around receptors has adopted a number of terms to refer to ligands having various effects on receptors. For clarity and consistency, the following definitions will be used throughout this patent document. To the extent that these definitions conflict with other definitions for these terms, the following definitions shall control:

AGONISTS shall mean materials (e.g., ligands, candidate compounds) that

In the first of the second of

25

activate the intracellular response when they bind to the receptor, or enhance GTP binding to membranes.

AMINO ACID ABBREVIATIONS used herein are set out in Table A:

| | | TABLE A | |
|----|---------------|---------|-----|
| 5 | ALANINE | ALA | A |
| | ARGININE | ARG | R |
| | ASPARAGINE | ASN | N |
| | ASPARTIC ACID | ASP | D |
| | CYSTEINE | CYS | С |
| 10 | GLUTAMIC ACID | GLU | E |
| | GLUTAMINE | GLN | Q |
| | GLYCINE | GLY | Ğ |
| | HISTIDINE | HIS | H |
| | ISOLEUCINE | ILE | I |
| 15 | LEUCINE | LEU | L |
| | LYSINE | LYS | K |
| | METHIONINE | MET | М . |
| | PHENYLALANINE | PHE | F |
| | PROLINE | PRO | P |
| 20 | SERINE | SER | S |
| | THREONINE | THR | T |
| | TRYPTOPHAN | TRP | W |
| | TYROSINE | TYR | Y |
| | VALINE | VAL | V |

PARTIAL AGONISTS shall mean materials (e.g., ligands, candidate compounds) that activate the intracellular response when they bind to the receptor to a lesser degree/extent than do agonists, or enhance GTP binding to membranes to a lesser degree/extent than do agonists.

ANTAGONIST shall mean materials (e.g., ligands, candidate compounds) that competitively bind to the receptor at the same site as the agonists but which do not activate the intracellular response initiated by the active form of the receptor, and can thereby inhibit the intracellular responses by agonists or partial agonists. ANTAGONISTS do not diminish the baseline intracellular response in the absence of an agonist or partial agonist.

CANDIDATE COMPOUND shall mean a molecule (for example, and not limitation,

div dan was ante m

15

AREN-0054 - 7 - PATENT

a chemical compound) that is amenable to a screening technique. Preferably, the phrase "candidate compound" does not include compounds which were publicly known to be compounds selected from the group consisting of inverse agonist, agonist or antagonist to a receptor, as previously determined by an indirect identification process ("indirectly identified compound"); more preferably, not including an indirectly identified compound which has previously been determined to have therapeutic efficacy in at least one mammal; and, most preferably, not including an indirectly identified compound which has previously been determined to have therapeutic utility in humans.

COMPOSITION means a material comprising at least one component; a "pharmaceutical composition" is an example of a composition.

COMPOUND EFFICACY shall mean a measurement of the ability of a compound to inhibit or stimulate receptor functionality, as opposed to receptor binding affinity. Exemplary means of detecting compound efficacy are disclosed in the Example section of this patent document.

CODON shall mean a grouping of three nucleotides (or equivalents to nucleotides) which generally comprise a nucleoside (adenosine (A), guanosine (G), cytidine (C), uridine (U) and thymidine (T)) coupled to a phosphate group and which, when translated, encodes an amino acid.

CONSTITUTIVELY ACTIVATED RECEPTOR shall mean a receptor subject to constitutive receptor activation. A constitutively activated receptor can be endogenous or non-endogenous.

CONSTITUTIVE RECEPTOR ACTIVATION shall mean stabilization of a receptor in the active state by means other than binding of the receptor with its endogenous

F

CONTACT or CONTACTING shall mean bringing at least two moieties together.

whether in an in vitro system or an in vivo system.

ligand or a chemical equivalent thereof.

DIRECTLY IDENTIFYING or DIRECTLY IDENTIFIED, in relationship to the

5 phrase "candidate compound", shall mean the screening of a candidate compound against a

constitutively activated receptor, preferably a constitutively activated orphan receptor, and

most preferably against a constitutively activated G protein-coupled cell surface orphan

receptor, and assessing the compound efficacy of such compound. This phrase is, under no

circumstances, to be interpreted or understood to be encompassed by or to encompass the

phrase "indirectly identifying" or "indirectly identified."

ENDOGENOUS shall mean a material that a mammal naturally produces. ENDOGENOUS in reference to, for example and not limitation, the term "receptor," shall mean that which is naturally produced by a mammal (for example, and not limitation, a human) or a virus. By contrast, the term NON-ENDOGENOUS in this context shall mean that which is not naturally produced by a mammal (for example, and not limitation, a human) or a virus. For example, and not limitation, a receptor which is not constitutively active in its endogenous form, but when manipulated becomes constitutively active, is most preferably referred to herein as a "non-endogenous, constitutively activated receptor." Both terms can be utilized to describe both "in vivo" and "in vitro" systems. For example, and not limitation, 20 in a screening approach, the endogenous or non-endogenous receptor may be in reference to an in vitro screening system. As a further example and not limitation, where the genome of a mammal has been manipulated to include a non-endogenous constitutively activated receptor, screening of a candidate compound by means of an in vivo system is viable.

AREN-0054 - 9 - PATENT

G PROTEIN COUPLED RECEPTOR FUSION PROTEIN and GPCR FUSION

PROTEIN, in the context of the invention disclosed herein, each mean a non-endogenous protein comprising an endogenous, constitutively activate GPCR or a non-endogenous, constitutively activated GPCR fused to at least one G protein, most preferably the alpha (α) subunit of such G protein (this being the subunit that binds GTP), with the G protein preferably being of the same type as the G protein that naturally couples with endogenous orphan GPCR. For example, and not limitation, in an endogenous state, if the G protein "Gsα" is the predominate G protein that couples with the GPCR, a GPCR Fusion Protein based upon the specific GPCR would be a non-endogenous protein comprising the GPCR fused to Gsα; in some circumstances, as will be set forth below, a non-predominant G protein can be fused to the GPCR. The G protein can be fused directly to the c-terminus of the constitutively active GPCR or there may be spacers between the two.

HOST CELL shall mean a cell capable of having a Plasmid and/or Vector incorporated therein. In the case of a prokaryotic Host Cell, a Plasmid is typically replicated as a autonomous molecule as the Host Cell replicates (generally, the Plasmid is thereafter isolated for introduction into a eukaryotic Host Cell); in the case of a eukaryotic Host Cell, a Plasmid is integrated into the cellular DNA of the Host Cell such that when the eukaryotic Host Cell replicates, the Plasmid replicates. Preferably, for the purposes of the invention disclosed herein, the Host Cell is eukaryotic, more preferably, mammalian, and most preferably selected from the group consisting of 293, 293T and COS-7 cells.

INDIRECTLY IDENTIFYING or INDIRECTLY IDENTIFIED means the traditional approach to the drug discovery process involving identification of an endogenous ligand specific for an endogenous receptor, screening of candidate compounds against the

receptor for determination of those which interfere and/or compete with the ligand-receptor interaction, and assessing the efficacy of the compound for affecting at least one second messenger pathway associated with the activated receptor.

INHIBIT or INHIBITING, in relationship to the term "response" shall mean that a response is decreased or prevented in the presence of a compound as opposed to in the absence of the compound.

INVERSE AGONISTS shall mean materials (e.g., ligand, candidate compound) which bind to either the endogenous form of the receptor or to the constitutively activated form of the receptor, and which inhibit the baseline intracellular response initiated by the active form of the receptor below the normal base level of activity which is observed in the absence of agonists or partial agonists, or decrease GTP binding to membranes. Preferably, the baseline intracellular response is inhibited in the presence of the inverse agonist by at least 30%, more preferably by at least 50%, and most preferably by at least 75%, as compared with the baseline response in the absence of the inverse agonist.

KNOWN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has been identified.

LIGAND shall mean an endogenous, naturally occurring molecule specific for an endogenous, naturally occurring receptor.

MUTANT or MUTATION in reference to an endogenous receptor's nucleic acid and/or amino acid sequence shall mean a specified change or changes to such endogenous sequences such that a mutated form of an endogenous, non-constitutively activated receptor evidences constitutive activation of the receptor. In terms of equivalents to specific sequences, a subsequent mutated form of a human receptor is considered to be equivalent to

10

15

20

a first mutation of the human receptor if (a) the level of constitutive activation of the subsequent mutated form of a human receptor is substantially the same as that evidenced by the first mutation of the receptor; and (b) the percent sequence (amino acid and/or nucleic acid) homology between the subsequent mutated form of the receptor and the first mutation of the receptor is at least about 80%, more preferably at least about 90% and most preferably at least 95%. Ideally, and owing to the fact that the most preferred cassettes disclosed herein for achieving constitutive activation includes a single amino acid and/or codon change between the endogenous and the non-endogenous forms of the GPCR, the percent sequence homology should be at least 98%.

NON-ORPHAN RECEPTOR shall mean an endogenous naturally occurring molecule specific for an endogenous naturally occurring ligand wherein the binding of a ligand to a receptor activates an intracellular signaling pathway.

ORPHAN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has not been identified or is not known.

PHARMACEUTICAL COMPOSITION shall mean a composition comprising at least one active ingredient, whereby the composition is amenable to investigation for a specified, efficacious outcome in a mammal (for example, and not limitation, a human). Those of ordinary skill in the art will understand and appreciate the techniques appropriate for determining whether an active ingredient has a desired efficacious outcome based upon the needs of the artisan.

PLASMID shall mean the combination of a Vector and cDNA. Generally, a Plasmid is introduced into a Host Cell for the purposes of replication and/or expression of the cDNA as a protein.

STIMULATE or STIMULATING, in relationship to the term "response" shall mean that a response is increased in the presence of a compound as opposed to in the absence of the compound.

VECTOR in reference to cDNA shall mean a circular DNA capable of incorporating

at least one cDNA and capable of incorporation into a Host Cell.

The order of the following sections is set forth for presentational efficiency and is not intended, nor should be construed, as a limitation on the disclosure or the claims to follow.

A. Introduction

The traditional study of receptors has always proceeded from the a priori assumption (historically based) that the endogenous ligand must first be identified before discovery could proceed to find antagonists and other molecules that could affect the receptor. Even in cases where an antagonist might have been known first, the search immediately extended to looking for the endogenous ligand. This mode of thinking has persisted in receptor research even after the discovery of constitutively activated receptors. What has not been heretofore recognized is that it is the active state of the receptor that is most useful for discovering agonists, partial agonists, and inverse agonists of the receptor. For those diseases which result from an overly active receptor or an under-active receptor, what is desired in a therapeutic drug is a compound which acts to diminish the active state of a receptor or enhance the activity of the receptor, respectively, not necessarily a drug which is an antagonist to the endogenous ligand. This is because a compound that reduces or enhances the activity of the active receptor state need not bind at the same site as the endogenous ligand. Thus, as taught by a method of this invention, any search for therapeutic compounds should start by screening compounds against the ligand-independent active state.

B. Identification of Human GPCRs

The efforts of the Human Genome project has led to the identification of a plethora of information regarding nucleic acid sequences located within the human genome; it has been the case in this endeavor that genetic sequence information has been made available without an understanding or recognition as to whether or not any particular genomic sequence does or may contain open-reading frame information that translate human proteins. Several methods of identifying nucleic acid sequences within the human genome are within the purview of those having ordinary skill in the art. For example, and not limitation, a variety of human GPCRs, disclosed herein, were discovered by reviewing the GenBankTM database, while other GPCRs were discovered by utilizing a nucleic acid sequence of a GPCR, previously sequenced, to conduct a BLASTTM search of the EST database. Table B, below, lists several endogenous GPCRs that we have discovered, along with a GPCR's respective homologous receptor.

TABLE B

| 15 | Disclosed Human Orphan GPCRs | Accession Number Identified | Open Reading Frame (Base Pairs) | Per Cent Homology To Designated GPCR | Reference To Homologous GPCR (Accession No.) |
|----|---------------------------------------|-----------------------------------|---------------------------------------|---|---|
| | hARE-3 | AL033379 | 1,260 bp | 52.3% LPA-R | U92642 |
| 20 | hARE-4 | AC006087 | 1,119 bp | 36% P2Y5 | AF000546 |
| | hARE-5 | AC006255 | 1,104 bp | 32% Oryzias latipes | D43633 |
| | hGPR27 | AA775870 | 1,128 bp | • | |
| | hARE-1 | AI090920 | 999 bp | 43% KIAA0001 | D13626 |
| | hARE-2 | AA359504 | 1,122 bp | 53% GPR27 | |
| 25 | hPPR1 | H67224 | 1,053 bp | 39% EBI1 | L31581 |
| | hG2A | AA754702 | 1,113 bp | 31% GPR4 | L36148 |

| AREN-0054 | | | - 14 - | | PATENT |
|-----------|---------|-------------|----------|----------------------------|---------------|
| | hRUP3 | AL035423 | 1,005 bp | 30% | 2133653 |
| | | | | Drosophila melanogaster | |
| | hRUP4 | AI307658 | 1,296 bp | 32% pNPGPR | NP_004876 |
| | | | | 28% and 29 % | AAC41276 |
| | | | | Zebra fish Ya | and |
| | | | | and Yb, | AAB94616 |
| | hRUP5 | AC005849 | 1 412 1 | respectively | |
| | IIICI 3 | AC003649 | 1,413 bp | 25% DEZ | Q99788 |
| | LDID | 1.000.50.50 | | 23% FMLPR | P21462 |
| _ | hRUP6 | AC005871 | 1,245 bp | 48% GPR66 | NP 006047 |
| 5 | hRUP7 | AC007922 | 1,173 bp | 43% H3R | AF140538 |
| | hCHN3 | EST 36581 | 1,113 bp | 53% GPR27 | |
| | hCHN4 | AA804531 | 1,077 bp | 32% thrombin | 4503637 |
| | hCHN6 | EST 2134670 | 1,503 bp | 36% edg-1 | NP_001391 |
| | hCHN8 | EST 764455 | 1,029 bp | 47% | D13626 |
| | | | • | KIAA0001 | |
| 10 | hCHN9 | EST 1541536 | 1,077 bp | 41% LTB4R | NM 000752 |
|] | hCHN10 | EST 1365839 | 1,055 bp | 35% P2Y | NM_002563 |

Receptor homology is useful in terms of gaining an appreciation of a role of the receptors within the human body. As the patent document progresses, we will disclose techniques for mutating these receptors to establish non-endogenous, constitutively activated versions of these receptors.

The techniques disclosed herein have also been applied to other human, orphan GPCRs known to the art, as will be apparent as the patent document progresses.

C. Receptor Screening

Screening candidate compounds against a non-endogenous, constitutively activated version of the human GPCRs disclosed herein allows for the direct identification of candidate compounds which act at this cell surface receptor, without requiring use of the receptor's endogenous ligand. By determining areas within the body where the endogenous version of human GPCRs disclosed herein is expressed and/or over-expressed, it is possible to determine related disease/disorder states which are associated with the expression and/or over-expression

of the receptor; such an approach is disclosed in this patent document.

With respect to creation of a mutation that may evidence constitutive activation of the human GPCR disclosed herein is based upon the distance from the proline residue at which is presumed to be located within TM6 of the GPCR; this algorithmic technique is disclosed in co-pending and commonly assigned patent document U.S. Serial Number 09/170,496, incorporated herein by reference. The algorithmic technique is not predicated upon traditional sequence "alignment" but rather a specified distance from the aforementioned TM6 proline residue. By mutating the amino acid residue located 16 amino acid residues from this residue (presumably located in the IC3 region of the receptor) to, most preferably, a lysine residue, such activation may be obtained. Other amino acid residues may be useful in the mutation at this position to achieve this objective.

D. Disease/Disorder Identification and/or Selection

The same

And the last well post to the source

As will be set forth in greater detail below, most preferably inverse agonists to the non-endogenous, constitutively activated GPCR can be identified by the methodologies of this invention. Such inverse agonists are ideal candidates as lead compounds in drug discovery programs for treating diseases related to this receptor. Because of the ability to directly identify inverse agonists to the GPCR, thereby allowing for the development of pharmaceutical compositions, a search for diseases and disorders associated with the GPCR is relevant. For example, scanning both diseased and normal tissue samples for the presence of the GPCR now becomes more than an academic exercise or one which might be pursued along the path of identifying an endogenous ligand to the specific GPCR. Tissue scans can be conducted across a broad range of healthy and diseased tissues. Such tissue scans provide a preferred first step in associating a specific receptor with a disease and/or disorder. See, for

example, co-pending application (docket number ARE-0050) for exemplary dot-blot and RT-PCR results of several of the GPCRs disclosed herein.

Preferably, the DNA sequence of the human GPCR is used to make a probe for (a) dot-blot analysis against tissue-mRNA, and/or (b) RT-PCR identification of the expression of the receptor in tissue samples. The presence of a receptor in a tissue source, or a diseased tissue, or the presence of the receptor at elevated concentrations in diseased tissue compared to a normal tissue, can be preferably utilized to identify a correlation with a treatment regimen, including but not limited to, a disease associated with that disease. Receptors can equally well be localized to regions of organs by this technique. Based on the known functions of the specific tissues to which the receptor is localized, the putative functional role of the receptor can be deduced.

E. Screening of Candidate Compounds

The first term are now and and the first term and the first first

1. Generic GPCR screening assay techniques

When a G protein receptor becomes constitutively active, it binds to a G protein (e.g., Gq, Gs, Gi, Gz, Go) and stimulates the binding of GTP to the G protein. The G protein then acts as a GTPase and slowly hydrolyzes the GTP to GDP, whereby the receptor, under normal conditions, becomes deactivated. However, constitutively activated receptors continue to exchange GDP to GTP. A non-hydrolyzable analog of GTP, [35S]GTPγS, can be used to monitor enhanced binding to membranes which express constitutively activated receptors. It is reported that [35S]GTPγS can be used to monitor G protein coupling to membranes in the absence and presence of ligand. An example of this monitoring, among other examples well-known and available to those in the art, was reported by Traynor and Nahorski in 1995. The preferred use of this assay system is for initial screening of candidate compounds because the

system is generically applicable to all G protein-coupled receptors regardless of the particular G protein that interacts with the intracellular domain of the receptor.

2. Specific GPCR screening assay techniques

Once candidate compounds are identified using the "generic" G protein-coupled receptor assay (*i.e.*, an assay to select compounds that are agonists, partial agonists, or inverse agonists), further screening to confirm that the compounds have interacted at the receptor site is preferred. For example, a compound identified by the "generic" assay may not bind to the receptor, but may instead merely "uncouple" the G protein from the intracellular domain.

a. Gs, Gz and Gi.

min min

- 10

20

Gs stimulates the enzyme adenylyl cyclase. Gi (and Gz and Go), on the other hand, inhibit this enzyme. Adenylyl cyclase catalyzes the conversion of ATP to cAMP; thus, constitutively activated GPCRs that couple the Gs protein are associated with increased cellular levels of cAMP. On the other hand, constitutively activated GPCRs that couple Gi (or Gz, Go) protein are associated with decreased cellular levels of cAMP. See, generally, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Thus, assays that detect cAMP can be utilized to determine if a candidate compound is, e.g., an inverse agonist to the receptor (i.e., such a compound would decrease the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a most preferred approach relies upon the use of anti-cAMP antibodies in an ELISA-based format. Another type of assay that can be utilized is a whole cell second messenger reporter system assay. Promoters on genes drive the expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or

10

15

20

transcription factor (CREB) that then binds to the promoter at specific sites called cAMP response elements and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing multiple cAMP response elements before the reporter gene, e.g., β -galactosidase or luciferase. Thus, a constitutively activated Gs-linked receptor causes the accumulation of cAMP that then activates the gene and expression of the reporter protein. The reporter protein such as β -galactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995).

b. Go and Gq.

Gq and Go are associated with activation of the enzyme phospholipase C, which in turn hydrolyzes the phospholipid PIP₂, releasing two intracellular messengers: diacycloglycerol (DAG) and inistol 1,4,5-triphoisphate (IP₃). Increased accumulation of IP₃ is associated with activation of Gq- and Go-associated receptors. *See, generally*, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Assays that detect IP₃ accumulation can be utilized to determine if a candidate compound is, *e.g.*, an inverse agonist to a Gq- or Go-associated receptor (*i.e.*, such a compound would decrease the levels of IP₃). Gq-associated receptors can also been examined using an AP1 reporter assay in that Gq-dependent phospholipase C causes activation of genes containing AP1 elements; thus, activated Gq-associated receptors will evidence an increase in the expression of such genes, whereby inverse agonists thereto will evidence a decrease in such expression, and agonists will evidence an increase in such expression. Commercially available assays for such detection are available.

3. GPCR Fusion Protein

T. T. T. T.

tan dan tan

The use of an endogenous, constitutively activate orphan GPCR or a non-endogenous, constitutively activated orphan GPCR, for use in screening of candidate compounds for the direct identification of inverse agonists, agonists and partial agonists provide an interesting screening challenge in that, by definition, the receptor is active even in the absence of an endogenous ligand bound thereto. Thus, in order to differentiate between, *e.g.*, the non-endogenous receptor in the presence of a candidate compound and the non-endogenous receptor in the absence of that compound, with an aim of such a differentiation to allow for an understanding as to whether such compound may be an inverse agonist, agonist, partial agonist or have no affect on such a receptor, it is preferred that an approach be utilized that can enhance such differentiation. A preferred approach is the use of a GPCR Fusion Protein.

Generally, once it is determined that a non-endogenous orphan GPCR has been constitutively activated using the assay techniques set forth above (as well as others), it is possible to determine the predominant G protein that couples with the endogenous GPCR.

15 Coupling of the G protein to the GPCR provides a signaling pathway that can be assessed.

Because it is most preferred that screening take place by use of a mammalian expression system, such a system will be expected to have endogenous G protein therein. Thus, by definition, in such a system, the non-endogenous, constitutively activated orphan GPCR will continuously signal. In this regard, it is preferred that this signal be enhanced such that in the presence of, e.g., an inverse agonist to the receptor, it is more likely that it will be able to more readily differentiate, particularly in the context of screening, between the receptor when it is contacted with the inverse agonist.

The GPCR Fusion Protein is intended to enhance the efficacy of G protein coupling

The state of the s

And the the

with the non-endogenous GPCR. The GPCR Fusion Protein is preferred for screening with a non-endogenous, constitutively activated GPCR because such an approach increases the signal that is most preferably utilized in such screening techniques. This is important in facilitating a significant "signal to noise" ratio; such a significant ratio is import preferred for the screening of candidate compounds as disclosed herein.

The construction of a construct useful for expression of a GPCR Fusion Protein is within the purview of those having ordinary skill in the art. Commercially available expression vectors and systems offer a variety of approaches that can fit the particular needs of an investigator. The criteria of importance for such a GPCR Fusion Protein construct is that the endogenous GPCR sequence and the G protein sequence both be in-frame (preferably, the sequence for the endogenous GPCR is upstream of the G protein sequence) and that the "stop" codon of the GPCR must be deleted or replaced such that upon expression of the GPCR, the G protein can also be expressed. The GPCR can be linked directly to the G protein, or there can be spacer residues between the two (preferably, no more than about 12, although this number can be readily ascertained by one of ordinary skill in the art). We have a preference (based upon convenience) of use of a spacer in that some restriction sites that are not used will, effectively, upon expression, become a spacer. Most preferably, the G protein that couples to the non-endogenous GPCR will have been identified prior to the creation of the GPCR Fusion Protein construct. Because there are only a few G proteins that have been identified, it is preferred that a construct comprising the sequence of the G protein (i.e., a universal G protein construct) be available for insertion of an endogenous GPCR sequence therein; this provides for efficiency in the context of large-scale screening of a variety of different endogenous GPCRs having different sequences.

20

As noted above, constitutively activated GPCRs that couple to Gi, Gz and Go are expected to inhibit the formation of cAMP making assays based upon these types of GPCRs challenging (*i.e.*, the cAMP signal decreases upon activation thus making the direct identification of, *e.g.*, inverse agonists (which would further decrease this signal), interesting). As will be disclosed herein, we have ascertained that for these types of receptors, it is possible to create a GPCR Fusion Protein that is not based upon the endogenous GPCR's endogenous G protein, in an effort to establish a viable cyclase-based assay. Thus, for example, a Gz coupled receptor such as H9, a GPCR Fusion Protein can be established that utilizes a Gs fusion protein – we believe that such a fusion construct, upon expression, "drives" or "forces" the non-endogenous GPCR to couple with, e.g., Gs rather than the "natural" Gz protein, such that a cyclase-based assay can be established. Thus, for Gi, Gz and Go coupled receptors, we prefer that that when a GPCR Fusion Protein is used and the assay is based upon detection of adenyl cyclase activity, that the fusion construct be established with Gs (or an equivalent G protein that stimulates the formation of the enzyme adenylyl cyclase).

15 F. Medicinal Chemistry

Generally, but not always, direct identification of candidate compounds is preferably conducted in conjunction with compounds generated via combinatorial chemistry techniques, whereby thousands of compounds are randomly prepared for such analysis. Generally, the results of such screening will be compounds having unique core structures; thereafter, these compounds are preferably subjected to additional chemical modification around a preferred core structure(s) to further enhance the medicinal properties thereof. Such techniques are known to those in the art and will not be addressed in detail in this patent document.

15

G. Pharmaceutical compositions

Candidate compounds selected for further development can be formulated into pharmaceutical compositions using techniques well known to those in the art. Suitable pharmaceutically-acceptable carriers are available to those in the art; for example, see Remington's Pharmaceutical Sciences, 16th Edition, 1980, Mack Publishing Co., (Oslo et al., eds.)

H. Other Utility

Although a preferred use of the non-endogenous versions the human GPCRs disclosed herein may be for the direct identification of candidate compounds as inverse agonists, agonists or partial agonists (preferably for use as pharmaceutical agents), these versions of human GPCRs can also be utilized in research settings. For example, in vitro and in vivo systems incorporating GPCRs can be utilized to further elucidate and understand the roles these receptors play in the human condition, both normal and diseased, as well as understanding the role of constitutive activation as it applies to understanding the signaling cascade. The value in non-endogenous human GPCRs is that their utility as a research tool is enhanced in that, because of their unique features, non-endogenous human GPCRs can be used to understand the role of these receptors in the human body before the endogenous ligand therefor is identified. Other uses of the disclosed receptors will become apparent to those in the art based upon, inter alia, a review of this patent document.

20 EXAMPLES

The following examples are presented for purposes of elucidation, and not limitation, of the present invention. While specific nucleic acid and amino acid sequences are disclosed herein, those of ordinary skill in the art are credited with the ability to make minor

modifications to these sequences while achieving the same or substantially similar results reported below. The traditional approach to application or understanding of sequence cassettes from one sequence to another (e.g. from rat receptor to human receptor or from human receptor A to human receptor B) is generally predicated upon sequence alignment techniques whereby the sequences are aligned in an effort to determine areas of commonality. The mutational approach disclosed herein does not rely upon this approach but is instead based upon an algorithmic approach and a positional distance from a conserved proline residue located within the TM6 region of human GPCRs. Once this approach is secured, those in the art are credited with the ability to make minor modifications thereto to achieve substantially the same results (i.e., constitutive activation) disclosed herein. Such modified approaches are considered within the purview of this disclosure

Example 1 ENDOGENOUS HUMAN GPCRS

1. Identification of Human GPCRs

Certain of the disclosed endogenous human GPCRs were identified based upon a review of the GenBank™ database information. While searching the database, the following cDNA clones were identified as evidenced below (Table C).

TABLE C

| 20 | Disclosed Human Orphan GPCRs | Accession Number | Complete DNA Sequence (Base Pairs) | Open Reading Frame (Base Pairs) | Nucleic Acid SEQ.ID. NO. | Amino Acid SEQ.ID. NO. |
|----|---------------------------------------|---------------------|--|---------------------------------------|-----------------------------------|---------------------------------|
| | hARE-3 | AL033379 | 111,389 bp | 1,260 bp | 1 | 2 |
| | hARE-4 | AC006087 | 226,925 bp | 1,119 bp | 3 | 4 |
| 25 | hARE-5 | AC006255 | 127,605 bp | 1,104 bp | 5 | 6 |
| | hRUP3 | AL035423 | 140,094 bp | 1,005 bp | 7 | 8 |

| AREN-0054 | - 24 - | | | | PATENT |
|-----------|----------|------------|----------|----|---------------|
| hRUP5 | AC005849 | 169,144 bp | 1,413 bp | 9 | 10 |
| hRUP6 | AC005871 | 218,807 bp | 1,245 bp | 11 | 12 |
| hRUP7 | AC007922 | 158,858 bp | 1,173 bp | 13 | 14 |

Other disclosed endogenous human GPCRs were identified by conducting a BLASTTM

5 search of EST database (dbest) using the following EST clones as query sequences. The following EST clones identified were then used as a probe to screen a human genomic library (Table D).

TABLE D

| 10 | Disclosed Human Orphan GPCRs | Query (Sequence) | EST Clone/ Accession No. Identified | Open Reading Frame (Base Pairs) | Nucleic Acid SEQ.ID.NO. | Amino Acid SEQ.ID.NO. |
|----|---------------------------------------|--|---|--|----------------------------|--------------------------|
| | hGPCR27 | Mouse GPCR27 | AA775870 | 1,125 bp | 17 | 18 |
| | hARE-1 | TDAG | 1689643 AI090920 | 999 bp | 19 | 20 |
| 15 | hARE-2 | GPCR27 | 68530 AA359504 | 1,122 bp | 21 | 22 |
| | hPPR1 | Bovine PPR1 | 238667 H67224 | 1,053 bp | 23 | 24 |
| | hG2A | Mouse 1179426 | See Example 2(a), below | 1,113 bp | 25 | 26 |
| | hCHN3 | N.A. | EST 36581 (full length) | 1,113 bp | 27 | 28 |
| | hCHN4 | TDAG | 1184934 AA804531 | 1,077 bp | 29 | 30 |
| 20 | hCHN6 | N.A. | EST 2134670 (full length) | 1,503 bp | 31 | 32 |
| | hCHN8 | KIAA0001 | EST 764455 | 1,029 bp | 33 | 34 |
| | hCHN 9 | 1365839 | EST 1541536 | 1,077 bp | 35 | 36 |
| | hCHN10 | Mouse EST 1365839 | Human 1365839 | 1,005 bp | 37 | 38 |
| | hRUP4 | N.A. | AI307658 | 1,296 bp | 39 | 40 |
| 25 | | N.A. = "not appear of ap | plicable". | | | |

2. Full Length Cloning

a. Human G2A

Mouse EST clone 1179426 was used to obtain a human genomic clone containing all

In the first of the section of the s

but three amino acid G2A coding sequences. The 5'of this coding sequence was obtained by using 5'RACE, and the template for PCR was Clontech's Human Spleen Marathon-Ready™ cDNA. The disclosed human G2A was amplified by PCR using the G2A cDNA specific primers for the first and second round PCR as shown in SEQ.ID.NO.: 41 and SEQ.ID.NO.:42 as follows:

- 5'-CTGTGTACAGCAGTTCGCAGAGTG-3' (SEQ.ID.NO.: 41; 1st round PCR)
- 5'-GAGTGCCAGGCAGAGCAGGTAGAC-3' (SEQ.ID.NO.: 42; second round PCR).

PCR was performed using Advantage GC Polymerase Kit (Clontech; manufacturing instructions will be followed), at 94°C for 30 sec followed by 5 cycles of 94°C for 5 sec and 72°C for 4 min; and 30 cycles of 94° for 5 sec and 70° for 4 min. An approximate 1.3 Kb PCR fragment was purified from agarose gel, digested with Hind III and Xba I and cloned into the expression vector pRC/CMV2 (Invitrogen). The cloned-insert was sequenced using the T7 Sequenase™ kit (USB Amersham; manufacturer instructions followed) and the sequence was compared with the presented sequence. Expression of the human G2A was detected by probing an RNA dot blot (Clontech; manufacturer instructions followed) with the P³²-labeled fragment.

b. CHN9

Sequencing of the EST clone 1541536 showed CHN9 to be a partial cDNA clone having only an initiation codon; *i.e.*, the termination codon was missing. When CHN9 was used to blast against data base (nr), the 3' sequence of CHN9 was 100% homologous to the 5' untranslated region of the leukotriene B4 receptor cDNA, which contained a termination codon in the frame with CHN9 coding sequence. To determine whether the 5' untranslated region of LTB4R cDNA was the 3' sequence of CHN9, PCR was performed using primers based upon the 5' sequence flanking the initiation codon found in CHN9 and

- the 3' sequence around the termination codon found in the LTB4R 5' untranslated region.

 The 5' primer sequence utilized was as follows:
- 5'-CCCGAATTCCTGCTTGCTCCCAGCTTGGCCC-3' (SEQ.ID.NO.: 43; sense) and
- 5'-TGTGGATCCTGCTGTCAAAGGTCCCATTCCGG-3' (SEQ.ID.NO.: 44; antisense).
- PCR was performed using thymus cDNA as a template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 uM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 65°C for 1 min and 72 °C for 1 min and 10 sec. A 1.1kb fragment consistent with the predicted size was obtained from PCR. This PCR fragment was subcloned into pCMV (see below) and sequenced (see, SEQ.ID.NO.: 35).

c. RUP 4

The full length RUP4 was cloned by RT-PCR with human brain cDNA (Clontech) as templates:

- 5'-TCACAATGCTAGGTGTGGTC-3' (SEQ.ID.NO.: 45; sense) and
- 5'-TGCATAGACAATGGGATTACAG-3' (SEQ.ID.NO.: 46; antisense).

PCR was performed using TaqPlus Precision™ polymerase (Stratagene; manufacturing instructions followed) by the following cycles: 94°C for 2 min; 94°C 30 sec; 55°C for 30 sec, 72°C for 45 sec, and 72°C for 10 min. Cycles 2 through 4 were repeated 30 times.

The PCR products were separated on a 1% agarose gel and a 500 bp PCR fragment was isolated and cloned into the pCRII-TOPO™ vector (Invitrogen) and sequenced using the T7 DNA Sequenase™ kit (Amsham) and the SP6/T7 primers (Stratagene). Sequence analysis revealed that the PCR fragment was indeed an alternatively spliced form of AI307658 having a continuous open reading frame with similarity to other GPCRs. The completed sequence of this PCR fragment was as follows:

- 10 Based on the above sequence, two sense oligonucleotide primer sets:
 - 5'-CTGCTTAGAAGAGTGGACCAG-3' (SEQ.ID.NO.: 48; oligo 1),
 - 5'-CTGTGCACCAGAAGATCTACAC-3' (SEQ.IDNO.: 49; oligo 2) and

two antisense oligonucleotide primer sets:

- 5'-CAAGGATGAAGGTGGTGTAGA-3' (SEQ.ID.NO.: 50; oligo 3)
- 15 5'-GTGTAGATCTTCTGGTGCACAGG-3' (SEQ.ID.NO.: 51; oligo 4)
 - were used for 3'- and 5'-RACE PCR with a human brain Marathon-Ready™ cDNA (Clontech, Cat# 7400-1) as template, according to manufacture's instructions. DNA fragments generated by the RACE PCR were cloned into the pCRII-TOPO™ vector (Invitrogen) and sequenced using the SP6/T7 primers (Stratagene) and some internal primers.
- The 3' RACE product contained a poly(A) tail and a completed open reading frame ending at a TAA stop codon. The 5' RACE product contained an incomplete 5' end; *i.e.*, the ATG initiation codon was not present.

Based on the new 5' sequence, oligo 3 and the following primer:

- 5'-GCAATGCAGGTCATAGTGAGC -3' (SEQ.ID.NO.: 52; oligo 5)
- were used for the second round of 5' race PCR and the PCR products were analyzed as above.

A third round of 5' race PCR was carried out utilizing antisense primers:

- 5'-TGGAGCATGGTGACGGGAATGCAGAAG-3' (SEQ.ID.NO.: 53; oligo 6) and
- 5'-GTGATGAGCAGGTCACTGAGCGCCAAG-3' (SEQ.ID.NO.: 54; oligo7).

The sequence of the 5' RACE PCR products revealed the presence of the initiation codon

ATG, and further round of 5' race PCR did not generate any more 5' sequence. The completed 5' sequence was confirmed by RT-PCR using sense primer

5'-GCAATGCAGGCGCTTAACATTAC-3' (SEQ.ID.NO.: 55; oligo 8)

and oligo 4 as primers and sequence analysis of the 650 bp PCR product generated from

human brain and heart cDNA templates (Clontech, Cat# 7404-1). The completed 3' sequence was confirmed by RT-PCR using oligo 2 and the following antisense primer:

5'-TTGGGTTACAATCTGAAGGGCA-3' (SEQ.ID.NO.:56; oligo 9)

and sequence analysis of the 670 bp PCR product generated from human brain and heart cDNA templates. (Clontech, Cat# 7404-1).

10 **d. RUP5**

The full length RUP5 was cloned by RT-PCR using a sense primer upstream from ATG, the initiation codon (SEQ.ID.NO.:57), and an antisense primer containing TCA as the stop codon (SEQ.ID.NO.:58), which had the following sequences:

5'-ACTCCGTGTCCAGCAGGACTCTG-3' (SEQ.ID.NO.: 57)

15 5'-TGCGTGTTCCTGGACCCTCACGTG-3' (SEQ.ID.NO.: 58)

and human peripheral leukocyte cDNA (Clontech) as a template. Advantage™ cDNA polymerase (Clontech) was used for the amplification in a 50ul reaction by the following cycle with step 2 through step 4 repeated 30 times: 94°C for 30 sec; 94° for 15 sec; 69° for 40 sec; 72°C for 3 min; and 72°C fro 6 min. A 1.4kb PCR fragment was isolated and cloned with the pCRII-TOPO™ vector (Invitrogen) and completely sequenced using the T7 DNA Sequenase™ kit (Amsham). *See*, SEQ.ID.NO.: 9.

e. RUP6

The full length RUP6 was cloned by RT-PCR using primers: 5'-CAGGCCTTGGATTTTAATGTCAGGGATGG-3' (SEQ.ID.NO.: 59) and

5'-GGAGAGTCAGCTCTGAAAGAATTCAGG-3' (SEQ.ID.NO.: 60);
and human thymus Marathon-Ready™ cDNA (Clontech) as a template. Advantage cDNA
polymerase (Clontech, according to manufacturer's instructions) was used for the
amplification in a 50ul reaction by the following cycle: 94°C for 30sec; 94°C for 5 sec; 66°C

for 40sec; 72°C for 2.5 sec and 72°C for 7 min. Cycles 2 through 4 were repeated 30 times.
A 1.3 Kb PCR fragment was isolated and cloned into the pCRII-TOPO™ vector (Invitrogen)
and completely sequenced (*see*, SEQ.ID.NO.: 11) using the ABI Big Dye Terminator™ kit
(P.E. Biosystem).

f. RUP7

The full length RUP7 was cloned by RT-PCR using primers:

5'-TGATGTGATGCCAGATACTAATAGCAC-3' (SEQ.ID.NO.: 61; sense) and

5'-CCTGATTCATTTAGGTGAGATTGAGAC-3' (SEQ.ID.NO.: 62; antisense)

and human peripheral leukocyte cDNA (Clontech) as a template. Advantage™ cDNA polymerase (Clontech) was used for the amplification in a 50 ul reaction by the following cycle with step 2 to step 4 repeated 30 times: 94°C for 2 minutes; 94°C for 15 seconds; 60°C for 20 seconds; 72°C for 2 minutes; 72°C for 10 minutes. A 1.25 Kb PCR fragment was isolated and cloned into the pCRII-TOPO™ vector (Invitrogen) and completely sequenced using the ABI Big Dye Terminator™ kit (P.E. Biosystem). See, SEQ.ID.NO.: 13.

3. Angiotensin II Type 1 Receptor ("AT1")

The endogenous human angiotensin II type 1 receptor ("AT1") was obtained by PCR using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 µM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 55°C for 1 min and 72 °C for 1.5 min. The 5' PCR primer contains a HindIII site with the sequence:

5'-CCCAAGCTTCCCCAGGTGTATTTGAT-3' (SEQ.ID.NO.: 63)

and the 3' primer contains a BamHI site with the following sequence:

5'-GTTGGATCCACATAATGCATTTTCTC-3' (SEQ.ID.NO.: 64).

The resulting 1.3 kb PCR fragment was digested with HindIII and BamHI and cloned into

HindIII-BamHI site of pCMV expression vector. The cDNA clone was fully sequenced.

Nucleic acid (SEQ.ID.NO.: 65) and amino acid (SEQ.ID.NO.: 66) sequences for human AT1

were thereafter determined and verified.

4. GPR38

To obtain GPR38, PCR was performed by combining two PCR fragments, using human genomic cDNA as template and rTth poymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25uM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition for each PCR reaction was 30 cycles of 94°C for 1 min, 62°C for 1 min and 72°C for 2 min.

The first fragment was amplified with the 5' PCR primer that contained an end site

with the following sequence:

5'-ACCATGGGCAGCCCCTGGAACGGCAGC-3' (SEQ.ID.NO.:67)

and a 3' primer having the following sequence:

5'-AGAACCACCACCAGCAGGACGGGGGGGTCTGCCGGTGG-3' (SEQ.ID.NO.:68).

The second PCR fragment was amplified with a 5' primer having the following sequence:

20 5'-GTCCGCGTCCTGCTGGTGGTGGTTCTGGCATTTATAATT-3' (SEQ.ID.NO.: 69)

and a 3' primer that contained a BamHI site and having the following sequence:

5'-CCTGGATCCTTATCCCATCGTCTTCACGTTAGC-3' (SEQ.ID.NO.: 70).

The two fragments were used as templates to amplify GPR38, using SEQ.ID.NO.: 67 and SEQ.ID.NO.: 70 as primers (using the above-noted cycle conditions). The resulting 1.44kb

PCR fragment was digested with BamHI and cloned into Blunt-BamHI site of pCMV expression vector.

5. MC4

To obtain MC4, PCR was performed using human genomic cDNA as template and rTth poymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25uM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition for each PCR reaction was 30 cycles of 94°C for 1 min, 54°C for 1min and 72°C for 1.5 min.

The 5' PCR contained an EcoRI site with the sequence:

5'-CTGGAATTCTCCTGCCAGCATGGTGA-3' (SEQ.ID.NO.: 71)

and the 3' primer contained a BamHI site with the sequence:

5'-GCAGGATCCTATATTGCGTGCTCTGTCCCC'-3 (SEQ.ID.NO.: 72).

The 1.0 kb PCR fragment was digest with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 73) and amino acid (SEQ.ID.NO.: 74) sequences for human MC4 were thereafter determined.

6. CCKB

To obtain CCKB, PCR was performed using human stomach cDNA as template and rTth poymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25uM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition for each PCR reaction was 30 cycles of 94°C for 1 min, 65°C for 1 min and 72°C for 1 min and 30 sec.

20 The 5' PCR contained a HindIII site with the sequence:

5'-CCGAAGCTTCGAGCTGAGTAAGGCGGCGGGCT-3' (SEQ.ID.NO.: 75)

and the 3' primer contained an EcoRI site with the sequence:

5'-GTGGAATTCATTTGCCCTGCCTCAACCCCCA-3 (SEQ.ID.NO.: 76).

The resulting 1.44 kb PCR fragment was digest with HindIII and EcoRI and cloned into

HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 77) and amino acid (SEQ.ID.NO.: 78) sequences for human CCKB were thereafter determined.

7. TDAG8

To obtain TDAG8, PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 56°C for 1 min and 72 °C for 1 min and 20 sec. The 5' PCR primer contained a HindIII site with the following sequence:

- 5'-TGCAAGCTTAAAAAGGAAAAAATGAACAGC-3' (SEQ.ID.NO.: 79)
- and the 3' primer contained a BamHI site with the following sequence:
 - 5'-TAAGGATCCCTTCCAAAACATCCTTG -3' (SEQ.ID.NO.: 80).

The resulting 1.1 kb PCR fragment was digested with HindIII and BamHI and cloned into HindIII-BamHI site of pCMV expression vector. Three resulting clones sequenced contained three potential polymorphisms involving changes of amino acid 43 from Pro to Ala, amino acid 97 from Lys to Asn and amino acid 130 from Ile to Phe. Nucleic acid (SEQ.ID.NO.: 81) and amino acid (SEQ.ID.NO.: 82) sequences for human TDAG8 were thereafter determined.

8. H9

To obtain H9, PCR was performed using pituitary cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 62°C for 1 min and 72°C for 2 min. The 5' PCR primer contained a HindIII site with the following sequence:

5'-GGAAAGCTTAACGATCCCCAGGAGCAACAT-3' (SEQ.ID.NO.:15)

and the 3' primer contained a BamHI site with the following sequence:

And the second of the second o

10

15

5'-CTGGGATCCTACGAGAGCATTTTTCACACAG-3' (SEQ.ID.NO.:16).

The resulting 1.9 kb PCR fragment was digested with HindIII and BamHI and cloned into HindIII-BamHI site of pCMV expression vector. H9 contained three potential polymorphisms involving changes of amino acid P320S, S493N and amino acid G448A. Nucleic acid (SEQ.ID.NO.: 139) and amino acid (SEQ.ID.NO.: 140) sequences for human H9 were thereafter determined and verified.

Example 2 PREPARATION OF NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED GPCRS

Those skilled in the art are credited with the ability to select techniques for mutation of a nucleic acid sequence. Presented below are approaches utilized to create non-endogenous versions of several of the human GPCRs disclosed above. The mutations disclosed below are based upon an algorithmic approach whereby the 16th amino acid (located in the IC3 region of the GPCR) from a conserved proline residue (located in the TM6 region of the GPCR, near the TM6/IC3 interface) is mutated, most preferably to a lysine amino acid residue.

1. Tranformer Site-Directed ™ Mutagenesis

Preparation of non-endogenous human GPCRs may be accomplished on human GPCRs using Transformer Site-Directed™ Mutagenesis Kit (Clontech) according to the manufacturer instructions. Two mutagenesis primers are utilized, most preferably a lysine mutagenesis oligonucleotide that creates the lysine mutation, and a selection marker oligonucleotide. For convenience, the codon mutation to be incorporated into the human GPCR is also noted, in standard form (Table E):

TABLE E

| | Receptor Identifier | Codon Mutation |
|----|---------------------|----------------|
| | hARE-3 | F313K |
| | hARE-4 | V233K |
| 5 | hARE-5 | A240K |
| | hGPCR14 | L257K |
| | hGPCR27 | C283K |
| | hARE-1 | E232K |
| | hARE-2 | G285K |
| 10 | hPPR1 | L239K |
| | hG2A | K232A |
| | hRUP3 | L224K |
| | hRUP5 | A236K |
| | hRUP6 | N267 K |
| 15 | hRUP7 | A302K |
| | hCHN4 | V236K |
| | hMC4 | A244K |
| | hCHN3 | S284K |
| | hCHN6 | L352K |
| 20 | hCHN8 | N235K |
| | hCHN9 | G223K |
| | hCHN10 | L231K |
| | hH9 | F236K |

The following GPCRs were mutated according with the above method using the

designated sequence primers (Table F).

TABLE F

PATENT

| | Receptor Identifier | Codon Mutation | Lysine Mutagenesis (SEQ.ID.NO.) 5'-3' orientation, mutation sequence underlined | Selection Marker (SEQ.ID.NO.) 5'-3' orientation |
|---|------------------------|-------------------|---|--|
| | hRUP4 | V272K | CAGGAAGAAG <u>AAA</u> CGAGC TGTCATTATGATGGTGACA GTG (83) | CACTGTCACCATCATAATG ACAGCTCGTTTCTTCC TG (84) |
| | hAT1 | see below | alternative approach; see below | alternative approach; see below |
| 5 | hGPR38 | V297K | GGCCACCGGCAGACC <u>AAA</u> C GCGTCCTGCTG (85) | CTCCTTCGGTCCTCCTATC GTTGTCAGAAGT (86) |
| | hCCKB | V332K | alternative approach; see below | alternative approach; see below |
| | hTDAG8 | 1225K | GGAAAAGAAGAGAATCAA <u>AAA</u> ACTACTTGTCAGCATC (87) | CTCCTTCGGTCCTCCTATC GTTGTCAGAAGT (88) |
| | hH9 | F236K | GCTGAGGTTCGCAAT <u>AAA</u> C TAACCATGTTTGTG (143) | CTCCTTCGGTCCTCCTATC GTTGTCAGAAGT (144) |
| | hMC4 | A244K | GCCAATATGAAGGGA <u>AAA</u> ATTACCTTGACCATC (137) | CTCCTTCGGTCCTCCTATC GTTGTCAGAAGT (138) |

The non-endogenous human GPCRs were then sequenced and the derived and verified nucleic acid and amino acid sequences are listed in the accompanying "Sequence Listing" appendix to this patent document, as summarized in Table G below:

TABLE G

| | Non Endogenous Human | Nucleic Acid Sequence Listing | Amino Acid Sequence |
|----|-----------------------------|-------------------------------|------------------------------|
| 15 | GPCR | | Listing |
| | hRUP4 | SEQ.ID.NO.: 127 | SEQ.ID.NO.: 128 |
| | (V272K) | | |
| | hAT1 | (see alternative approaches | (see alternative approaches, |
| | (see alternative approaches | below) | below) |
| 20 | below) | | |
| | hGPR38 | SEQ.ID.NO.: 129 | SEQ.ID.NO.: 130 |
| | (V297K) | | |
| | hCCKB | SEQ.ID.NO.: 131 | SEQ.ID.NO.: 132 |
| | (V332K) | | |
| 25 | HTDAG8 | SEQ.ID.NO.: 133 | SEQ.ID.NO.: 134 |
| | (I225K) | | |
| | hH9 | SEQ.ID.NO.: 141 | SEQ.ID.NO.: 142 |
| | (F236K) | | |
| | hMC4 | SEQ.ID.NO.: 135 | SEQ.ID.NO.: 136 |
| 30 | (A244K) | | |
| | | | |

2. Alternative Approaches For Creation of Non-Endogenous Human GPCRs

a. AT1

1. F239K Mutation

Preparation of a non-endogenous, constitutively activated human AT1 receptor was accomplished by creating an F239K mutation (see, SEQ.ID.NO.: 89 for nucleic acid sequence, and SEQ.ID.NO.: 90 for amino acid sequence). Mutagenesis was performed using Transformer Site-Directed Mutagenesis™ Kit (Clontech) according to the to manufacturer's instructions. The two mutagenesis primers were used, a lysine mutagenesis oligonucleotide (SEQ.ID.NO.: 91) and a selection marker oligonucleotide (SEQ.ID.NO.: 92), which had the following sequences:

5'-CCAAGAAATGATGATATTAAAAAGATAATTATGGC-3' (SEQ.ID.NO.: 91)

- 5'-CCAAGAAATGATGATATTAAAAAGATAATTATGGC-3' (SEQ.ID.NO.: 91)
 5'-CTCCTTCGGTCCTCCTATCGTTGTCAGAAGT-3' (SEQ.ID.NO.: 92),
- 15 respectively.

5

2. N111A Mutation

Preparation of a non-endogenous human AT1 receptor was also accomplished by creating an N111A mutation (see, SEQ.ID.NO.:93 for nucleic acid sequence, and SEQ.ID.NO.: 94 for amino acid sequence). Two PCR reactions were performed using pfu polymerase (Stratagene) with the buffer system provided by the manufacturer, supplemented with 10% DMSO, 0.25 μM of each primer, and 0.5 mM of each 4 nucleotides. The 5' PCR sense primer used had the following sequence: 5'-CCCAAGCTTCCCCAGGTGTATTTGAT-3' (SEQ.ID.NO.: 95)

25 and the antisense primer had the following sequence:

٠...

find the find

the second control of the second control of

The resulting 400 bp PCR fragment was digested with HindIII site and subcloned into HindIII-SmaI site of pCMV vector (5' construct). The 3' PCR sense primer used had the following sequence:

- 5'-CTGTACGCTAGTGTTTTCTACTCACGTGTCTCAGCATTGAT-3' (SEQ.ID.NO.: 97) and the antisense primer had the following sequence:
 - 5'-GTTGGATCCACATAATGCATTTTCTC-3' (SEQ.ID.NO.: 98)

The resulting 880 bp PCR fragment was digested with BamHI and inserted into Pst (blunted by T4 polymerase) and BamHI site of 5' construct to generated the full length N111A construct. The cycle condition was 25 cycles of 94°C for 1 min, 60°C for 1min and 72 °C for 1 min (5' PCR) or 1.5 min (3' PCR).

AT2K255IC3 Mutation 3.

Preparation of a non-endogenous, constitutively activated human AT1 was accomplished by creating an AT2K255IC3 "domain swap" mutation (see, SEQ.ID.NO.:99 for nucleic acid sequence, and SEQ.ID.NO.: 100 for amino acid sequence). Restriction 15 sites flanking IC3 of AT1 were generated to facilitate replacement of the IC3 with corresponding IC3 from angiotensin II type 2 receptor (AT2). This was accomplished by performing two PCR reactions. A 5' PCR fragment (Fragment A) encoded from the 5' untranslated region to the beginning of IC3 was generated by utilizing SEQ.ID.NO.: 63 as sense primer and the following sequence: 20

- 5'-TCCGAATTCCAAAATAACTTGTAAGAATGATCAGAAA-3' (SEQ.ID.NO.: 101)
- as antisense primer. A 3' PCR fragment (Fragment B) encoding from the end of IC3 to the
- 3' untranslated region was generated by using the following sequence:
- 5'-AGATCTTAAGAAGATAATTATGGCAATTGTGCT-3' (SEQ.ID.NO.: 102)

as sense primer and SEQ.ID.NO.: 64 as antisense primer. The PCR condition was 30 cycles of 94°C for 1 min, 55°C for 1min and 72°C for 1.5 min using endogenous AT1 cDNA clone as template and pfu polymerase (Stratagene), with the buffer systems provided by the manufacturer, supplemented with 10% DMSO, 0.25 µM of each primer, and 0.5 mM of each 4 nucleotides. Fragment A (720 bp) was digested with HindIII and EcoRI and subcloned. Fragment B was digested with BamHI and subcloned into pCMV vector with an EcoRI site 5' to the cloned PCR fragment.

The DNA fragment (Fragment C) encoding IC3 of AT2 with a L255K mutation and containing an EcoRI cohesive end at 5' and a AfIII cohesive end at 3', was generated by annealing 2 synthetic oligonucleotides having the following sequences:

5'AATTCGAAAACACTTACTGAAGACGAATAGCTATGGGAAGAACAGGATAACCCGTGACCAA G-3' (sense; SEQ.ID.NO.: 103)

5'TTAACTTGGTCACGGGTTATCCTGTTCTTCCCATAGCTATTCGTCTTCAGT AAGTGTTTTCG-3' (antisense; SEQ.ID.NO.: 104). 15

Fragment C was inserted in front of Fragment B through EcoRI and AfIII site. The resulting clone was then ligated with the Fragment A through the EcoRI site to generate AT1 with AT2K255IC3.

4. A243+ Mutation

Preparation of a non-endogenous human AT1 receptor was also accomplished by creating an A243+ mutation (see, SEQ.ID.NO.: 105 for nucleic acid sequence, and SEQ.ID.NO.: 106 for amino acid sequence). An A243+ mutation was constructed using the following PCR based strategy: Two PCR reactions was performed using pfu polymerase (Stratagene) with the buffer system provided by the manufacturer supplemented with 10% 25 DMSO, 0.25 µM of each primer, and 0.5 mM of each 4 nucleotides. The 5' PCR sense primer utilized had the following sequence:

5'-CCCAAGCTTCCCCAGGTGTATTTGAT-3' (SEQ.ID.NO.: 107)

and the antisense primer had the following sequence:

- 5'-AAGCACAATTGCTGCATAATTATCTTAAAAATATCATC-3' (SEQ.ID.NO.: 108).
- 5 The 3' PCR sense primer utilized had the following sequence:
 - 5'-AAGATAATTATGGCAGCAATTGTGCTTTTCTTTTTCTTT-3' (SEQ.ID.NO.: 109)

containing the Ala insertion and antisense primer:

5'-GTTGGATCCACATAATGCATTTTCTC-3'(SEQ.ID.NO.: 110).

The cycle condition was 25 cycles of 94°C for 1 min, 54°C for 1 min and 72 °C for 1.5 min.

An aliquot of the 5' and 3' PCR were then used as co-template to perform secondary PCR using the 5' PCR sense primer and 3' PCR antisense primer. The PCR condition was the same as primary PCR except the extention time was 2.5 min. The resulting PCR fragment was digested with HindIII and BamHI and subcloned into pCMV vector. (See,

SEQ.ID.NO.: 105)

15

20

4. CCKB

Preparation of the non-endogenous, constitutively activated human CCKB receptor was accomplished by creating a V322K mutation (see, SEQ.ID.NO.: 111 for nucleic acid sequence and SEQ.ID.NO.: 112 for amino acid sequence). Mutagenesis was performed by PCR via amplification using the wildtype CCKB from Example 1.

The first PCR fragment (1kb) was amplified by using SEQ.ID.NO.: 75 and an antisense primer comprising a V322K mutation:

5'-CAGCAGCATGCGCTTCACGCGCTTCTTAGCCCAG-3' (SEQ.ID.NO.: 113).

The second PCR fragment (0.44kb) was amplified by using a sense primer comprising the V322K mutation:

List have some once the most than the man one or one of the list has been some or one.

5'-AGAAGCGCGTGAAGCGCATGCTGCTGGTGATCGTT-3' (SEQ.ID.NO.: 114) and SEQ.ID.NO.: 76.

The two resulting PCR fragments were then used as template for amplifying CCKB comprising V332K, using SEQ.ID.NO.: 75 and SEQ.ID.NO.: 76 and the above-noted system and conditions. The resulting 1.44kb PCR fragment containing the V332K mutation was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. (*See*, SEQ.ID.NO.: 111).

3. QuikChange™ Site-Directed™ Mutagenesis

Preparation of non-endogenous human GPCRs can also be accomplished by using

QuikChangeTM Site-DirectedTM Mutagenesis Kit (Stratagene, according to manufacturer's instructions). Endogenous GPCR is preferably used as a template and two mutagenesis primers utilized, as well as, most preferably, a lysine mutagenesis oligonucleotide and a selection marker oligonucleotide (included in kit). For convenience, the codon mutation incorporated into the human GPCR and the respective oligonucleotides are noted, in standard form (Table H):

TABLE H

| | Receptor Identifier | Codon Mutation | Lysine Mutagenesis (SEQ.ID.NO.) 5'-3' orientation, mutation underlined | Selection Marker (SEQ.ID.NO.) 5'-3' orientation |
|---|------------------------|-------------------|--|---|
| | hCHN3 | S284K | ATGGAGAAAAGAATC <u>AAA</u> AGAA TGTTCTATATA (115) | TATATAGAACATTCTTTT GATTCTTTTCTCCAT (116) |
| | hCHN6 | L352K | CGCTCTCTGGCCTTG <u>AAG</u> CGCAC GCTCAGC (117) | GCTGAGCGTGCGCTTCA AGGCCAGAGAGCG (118) |
| 5 | hCHN8 | N235K | CCCAGGAAAAAGGTG <u>AAA</u> GTCA AAGTTTTC (119) | GAAAACTTTGACTTTCAC CTTTTTCCTGGG (120) |
| | hCHN9 | G223K | GGGGCGCGGTG <u>AAA</u> CGGCTGG TGAGC (121) | GCTCACCAGCCGTTTCA CCCGCGCCCC (122) |
| | hCHN10 | L231K | CCCCTTGA <u>AAA</u> GCCTAAGAACTT GGTCATC (123) | GATGACCAAGTTCTTAG GCTTTTCAAGGGG (124) |

Example 3 RECEPTOR EXPRESSION

Although a variety of cells are available to the art for the expression of proteins, it is most preferred that mammalian cells be utilized. The primary reason for this is predicated upon practicalities, *i.e.*, utilization of, *e.g.*, yeast cells for the expression of a GPCR, while possible, introduces into the protocol a non-mammalian cell which may not (indeed, in the case of yeast, does not) include the receptor-coupling, genetic-mechanism and secretary pathways that have evolved for mammalian systems – thus, results obtained in non-mammalian cells, while of potential use, are not as preferred as that obtained from mammalian cells. Of the mammalian cells, COS-7, 293 and 293T cells are particularly preferred, although the specific mammalian cell utilized can be predicated upon the particular needs of the artisan.

On day one, 1X10⁷ 293T cells per 150mm plate were plated out. On day two, two reaction tubes were prepared (the proportions to follow for each tube are per plate): tube A was prepared by mixing 20µg DNA (e.g., pCMV vector; pCMV vector with receptor cDNA, etc.) in 1.2ml serum free DMEM (Irvine Scientific, Irvine, CA); tube B was

prepared by mixing 120μl lipofectamine (Gibco BRL) in 1.2ml serum free DMEM. Tubes A and B were admixed by inversions (several times), followed by incubation at room temperature for 30-45min. The admixture is referred to as the "transfection mixture".

Plated 293T cells were washed with 1XPBS, followed by addition of 10ml serum free

DMEM. 2.4ml of the transfection mixture were added to the cells, followed by incubation for 4hrs at 37°C/5% CO₂. The transfection mixture was removed by aspiration, followed by the addition of 25ml of DMEM/10% Fetal Bovine Serum. Cells were incubated at 37°C/5% CO₂. After 72hr incubation, cells were harvested and utilized for analysis.

Example 4 ASSAYS FOR DETERMINATION OF CONSTITUTIVE ACTIVITY OF NON-ENDOGENOUS GPCRS

A variety of approaches are available for assessment of constitutive activity of the non-endogenous human GPCRs. The following are illustrative; those of ordinary skill in the art are credited with the ability to determine those techniques that are preferentially beneficial for the needs of the artisan.

1. Membrane Binding Assays: [35S]GTPγS Assay

When a G protein-coupled receptor is in its active state, either as a result of ligand binding or constitutive activation, the receptor couples to a G protein and stimulates the release of GDP and subsequent binding of GTP to the G protein. The alpha subunit of the G protein-receptor complex acts as a GTPase and slowly hydrolyzes the GTP to GDP, at which point the receptor normally is deactivated. Constitutively activated receptors continue to exchange GDP for GTP. The non-hydrolyzable GTP analog, [35S]GTPγS, can be utilized to demonstrate enhanced binding of [35S]GTPγS to membranes expressing constitutively activated receptors. The advantage of using [35S]GTPγS binding to measure constitutive

activation is that: (a) it is generically applicable to all G protein-coupled receptors; (b) it is proximal at the membrane surface making it less likely to pick-up molecules which affect the intracellular cascade.

The assay utilizes the ability of G protein coupled receptors to stimulate [35S]GTPγS binding to membranes expressing the relevant receptors. The assay can, therefore, be used in the direct identification method to screen candidate compounds to known, orphan and constitutively activated G protein-coupled receptors. The assay is generic and has application to drug discovery at all G protein-coupled receptors.

The [35S]GTPγS assay can be incubated in 20 mM HEPES and between 1 and about 20mM MgCl₂ (this amount can be adjusted for optimization of results, although 20mM is preferred) pH 7.4, binding buffer with between about 0.3 and about 1.2 nM [35S]GTPγS (this amount can be adjusted for optimization of results, although 1.2 is preferred) and 12.5 to 75 μg membrane protein (e.g, COS-7 cells expressing the receptor; this amount can be adjusted for optimization, although 75μg is preferred) and 1 μM GDP (this amount can be changed for optimization) for 1 hour. Wheatgerm agglutinin beads (25 μl; Amersham) should then be added and the mixture incubated for another 30 minutes at room temperature. The tubes are then centrifuged at 1500 x g for 5 minutes at room temperature and then counted in a scintillation counter.

and the second control of the second control

10

15

A less costly but equally applicable alternative has been identified which also meets
the needs of large scale screening. Flash platesTM and WallacTM scintistrips may be utilized
to format a high throughput [35S]GTPγS binding assay. Furthermore, using this technique,
the assay can be utilized for known GPCRs to simultaneously monitor tritiated ligand binding
to the receptor at the same time as monitoring the efficacy via [35S]GTPγS binding. This is

20

possible because the Wallac beta counter can switch energy windows to look at both tritium and ³⁵S-labeled probes. This assay may also be used to detect other types of membrane activation events resulting in receptor activation. For example, the assay may be used to monitor ³²P phosphorylation of a variety of receptors (both G protein coupled and tyrosine kinase receptors). When the membranes are centrifuged to the bottom of the well, the bound [³⁵S]GTPγS or the ³²P-phosphorylated receptor will activate the scintillant which is coated of the wells. Scinti[®] strips (Wallac) have been used to demonstrate this principle. In addition, the assay also has utility for measuring ligand binding to receptors using radioactively labeled ligands. In a similar manner, when the radiolabeled bound ligand is centrifuged to the bottom of the well, the scintistrip label comes into proximity with the radiolabeled ligand resulting in activation and detection.

2. Adenylyl Cyclase

A Flash PlateTM Adenylyl Cyclase kit (New England Nuclear; Cat. No. SMP004A) designed for cell-based assays can be modified for use with crude plasma membranes. The Flash Plate wells contain a scintillant coating which also contains a specific antibody recognizing cAMP. The cAMP generated in the wells was quantitated by a direct competition for binding of radioactive cAMP tracer to the cAMP antibody. The following serves as a brief protocol for the measurement of changes in cAMP levels in membranes that express the receptors.

Transfected cells are harvested approximately three days after transfection.

Membranes were prepared by homogenization of suspended cells in buffer containing 20mM

HEPES, pH 7.4 and 10mM MgCl₂. Homogenization is performed on ice using a Brinkman

Polytron[™] for approximately 10 seconds. The resulting homogenate is centrifuged at 49,000

X g for 15 minutes at 4°C. The resulting pellet is then resuspended in buffer containing 20mM HEPES, pH 7.4 and 0.1 mM EDTA, homogenized for 10 seconds, followed by centrifugation at 49,000 X g for 15 minutes at 4°C. The resulting pellet can be stored at -80°C until utilized. On the day of measurement, the membrane pellet is slowly thawed at room temperature, resuspended in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCL₂ (these amounts can be optimized, although the values listed herein are preferred), to yield a final protein concentration of 0.60mg/ml (the resuspended membranes were placed on ice until use).

cAMP standards and Detection Buffer (comprising 2 μ Ci of tracer [125] cAMP (100 μ l] to 11 ml Detection Buffer) are prepared and maintained in accordance with the manufacturer's instructions. Assay Buffer is prepared fresh for screening and contained 20mM HEPES, pH 7.4, 10mM MgCl₂, 20mM (Sigma), 0.1 units/ml creatine phosphokinase (Sigma), 50 μ M GTP (Sigma), and 0.2 mM ATP (Sigma); Assay Buffer can be stored on ice until utilized. The assay is initiated by addition of 50ul of assay buffer followed by addition of 50ul of membrane suspension to the NEN Flash Plate. The resultant assay mixture is incubated for 60 minutes at room temperature followed by addition of 100ul of detection buffer. Plates are then incubated an additional 2-4 hours followed by counting in a Wallac MicroBetaTM scintillation counter. Values of cAMP/well are extrapolated from a standard cAMP curve that is contained within each assay plate.

C. Reporter-Based Assays

1. CREB Reporter Assay (Gs-associated receptors)

A method to detect Gs stimulation depends on the known property of the transcription factor CREB, which is activated in a cAMP-dependent manner. A PathDetect™ CREB trans-

15

20

Reporting System (Stratagene, Catalogue # 219010) can utilized to assay for Gs coupled activity in 293 or 293T cells. Cells are transfected with the plasmids components of this above system and the indicated expression plasmid encoding endogenous or mutant receptor using a Mammalian Transfection Kit (Stratagene, Catalogue #200285) according to the 5 manufacturer's instructions. Briefly, 400 ng pFR-Luc (luciferase reporter plasmid containing Gal4 recognition sequences), 40 ng pFA2-CREB (Gal4-CREB fusion protein containing the Gal4 DNA-binding domain), 80 ng pCMV-receptor expression plasmid (comprising the receptor) and 20 ng CMV-SEAP (secreted alkaline phosphatase expression plasmid; alkaline phosphatase activity is measured in the media of transfected cells to control for variations in transfection efficiency between samples) are combined in a calcium phosphate precipitate as per the Kit's instructions. Half of the precipitate is equally distributed over 3 wells in a 96well plate, kept on the cells overnight, and replaced with fresh medium the following morning. Forty-eight (48) hr after the start of the transfection, cells are treated and assayed for, e.g., luciferase activity

AP1 reporter assay (Gq-associated receptors) 2.

A method to detect Gq stimulation depends on the known property of Gq-dependent phospholipase C to cause the activation of genes containing AP1 elements in their promoter. A Pathdetect™ AP-1 cis-Reporting System (Stratagene, Catalogue # 219073) can be utilized following the protocol set forth above with respect to the CREB reporter assay, except that the components of the calcium phosphate precipitate were 410 ng pAP1-Luc, 80 ng pCMVreceptor expression plasmid, and 20 ng CMV-SEAP.

CRE-LUC Reporter Assay 3.

293 and 293T cells are plated-out on 96 well plates at a density of 2 x 10⁴ cells per

15

20

counter (Wallac).

well and were transfected using Lipofectamine Reagent (BRL) the following day according to manufacturer instructions. A DNA/lipid mixture is prepared for each 6-well transfection as follows: 260ng of plasmid DNA in 100µl of DMEM were gently mixed with 2µl of lipid in 100µl of DMEM (the 260ng of plasmid DNA consisted of 200ng of a 8xCRE-Luc reporter plasmid (see below and Figure 1 for a representation of a portion of the plasmid), 50ng of pCMV comprising endogenous receptor or non-endogenous receptor or pCMV alone, and 10ng of a GPRS expression plasmid (GPRS in pcDNA3 (Invitrogen)). The 8XCRE-Luc reporter plasmid was prepared as follows: vector SRIF- β -gal was obtained by cloning the rat somatostatin promoter (-71/+51) at BglV-HindIII site in the pβgal-Basic Vector (Clontech). Eight (8) copies of cAMP response element were obtained by PCR from an adenovirus template AdpCF126CCRE8 (see, 7 Human Gene Therapy 1883 (1996)) and cloned into the SRIF-β-gal vector at the Kpn-BglV site, resulting in the 8xCRE-β-gal reporter vector. The 8xCRE-Luc reporter plasmid was generated by replacing the beta-galactosidase gene in the 8xCRE-β-gal reporter vector with the luciferase gene obtained from the pGL3-basic vector (Promega) at the HindIII-BamHI site. Following 30 min. incubation at room temperature, the DNA/lipid mixture was diluted with 400 µl of DMEM and 100µl of the diluted mixture was added to each well. 100 µl of DMEM with 10% FCS were added to each well after a 4hr incubation in a cell culture incubator. The following day the transfected cells were changed with 200 µl/well of DMEM with 10% FCS. Eight (8) hours later, the wells were changed to $100\,\mu l$ /well of DMEM without phenol red, after one wash with PBS. Luciferase activity were measured the next day using the LucLite™ reporter gene assay kit (Packard) following manufacturer instructions and read on a 1450 MicroBeta™ scintillation and luminescence

15

4. SRF-LUC Reporter Assay

One method to detect Gq stimulation depends on the known property of Gq-dependent phospholipase C to cause the activation of genes containing serum response factors in their promoter. A Pathdetect™ SRF-Luc-Reporting System (Stratagene) can be utilized to assay for Gq coupled activity in, e.g., COS7 cells. Cells are transfected with the plasmid components of the system and the indicated expression plasmid encoding endogenous or nonendogenous GPCR using a Mammalian Transfection™ Kit (Stratagene, Catalogue #200285) according to the manufacturer's instructions. Briefly, 410 ng SRF-Luc, 80 ng pCMV-receptor expression plasmid and 20 ng CMV-SEAP (secreted alkaline phosphatase expression plasmid; alkaline phosphatase activity is measured in the media of transfected cells to control for variations in transfection efficiency between samples) are combined in a calcium phosphate precipitate as per the manufacturer's instructions. Half of the precipitate is equally distributed over 3 wells in a 96-well plate, kept on the cells in a serum free media for 24 hours. The last 5 hours the cells are incubated with 1 µM Angiotensin, where indicated. Cells are then lysed and assayed for luciferase activity using a Luclite™ Kit (Packard, Cat. #6016911) and "Trilux 1450 Microbeta" liquid scintillation and luminescence counter (Wallac) as per the manufacturer's instructions. The data can be analyzed using GraphPad Prism™ 2.0a (GraphPad Software Inc.).

5. Intracellular IP, Accumulation Assay

On day 1, cells comprising the receptors (endogenous and/or non-endogenous) can be plated onto 24 well plates, usually 1x10⁵ cells/well (although his umber can be optimized. On day 2 cells can be transfected by firstly mixing 0.25ug DNA in 50 ul serum free DMEM/well and 2 ul lipofectamine in 50 µl serumfree DMEM/well. The solutions

are gently mixed and incubated for 15-30 min at room temperature. Cells are washed with 0.5 ml PBS and 400 µl of serum free media is mixed with the transfection media and added to the cells. The cells are then incubated for 3-4 hrs at 37°C/5%CO₂ and then the transfection media is removed and replaced with 1ml/well of regular growth media. On day 3 the cells are labeled with ³H-myo-inositol. Briefly, the media is removed and the cells are washed with 0.5 ml PBS. Then 0.5 ml inositol-free/serum free media (GIBCO BRL) is added/well with 0.25 μ Ci of ³H-myo-inositol / well and the cells are incubated for 16-18 hrs o/n at 37°C/5%CO₂. On Day 4 the cells are washed with 0.5 ml PBS and 0.45 ml of assay medium is added containing inositol-free/serum free media 10 μ M pargyline 10 mM lithium chloride or 0.4 ml of assay medium and 50 ul of 10x ketanserin (ket) to final concentration of $10\mu M$. The cells are then incubated for 30 min at $37^{\circ}C$. The cells are then washed with 0.5 ml PBSand 200 ul of fresh/icecold stop solution (1M KOH; 18 mM Na-borate; 3.8 mM EDTA) is added/well. The solution is kept on ice for 5-10 min or until cells were lysed and then neutralized by 200 μ l of fresh/ice cold neutralization sol. (7.5 % HCL). The lysate is then transferred into 1.5 ml eppendorf tubes and 1 ml of chloroform/methanol (1:2) is added/tube. The solution is vortexed for 15 sec and the upper phase is applied to a Biorad AG1-X8™ anion exchange resin (100-200 mesh). Firstly, the resin is washed with water at 1:1.25 W/V and 0.9 ml of upper phase is loaded onto the column. The column is washed with 10 mls of 5 mM myo-inositol and 10 ml of 5 mM Na-borate/60mM Na-formate. The inositol tris phosphates are eluted into scintillation vials containing 10 ml of scintillation cocktail with 2 ml of 0.1 M formic acid/ 1 M ammonium formate. The columns are regenerated by washing with 10 ml of 0.1 M formic acid/3M ammonium formate and rinsed twice with dd H₂O and stored at 4°C in water.

10

15

AREN-0054 - 50 - PATENT

Exemplary results are presented below in Table I:

TABLE I

| | Receptor | Mutation | Assay Utilized | Signal Generated: Endogenous Version (Relative Light Units) | Signal Generated: Non- Endogenous Version (Relative Light Units) | Percent Difference |
|---|--------------|----------------|-------------------------|---|--|-----------------------|
| | hAT1 | F239K | SRF-LUC | 34 | 137 | 75%1 |
| | | AT2K255IC3 | SRF-LUC | 34 | 127 | 73%1 |
| 5 | hTDAG8 | I225K | CRE-LUC (293 cells) | 2,715 | 14,440 | 81%† |
| | | I225K | CRE-LUC (293T cells) | 65,681 | 185,636 | 65%↑ |
| | hH9 hCCKB | F236K V332K | CRE-LUC CRE-LUC | 1,887 785 | 6,096 3,223 | 69%† 76%† |

C. CELL-BASED DETECTION ASSAY (EXAMPLE -TDAG8)

293 cells were plated-out on 150mm plates at a density of 1.3 x 10⁷ cells per plate, and were transfected using 12ug of the respective DNA and 60ul of Lipofectamine Reagent (BRL) per plate. The transfected cells were grown in media containing serum for an assay performed 24 hours post-transfection. For detection assay performed 48 hours post-transfection (assay comparing serum and serum-free media; see Figure 3), the initial media was changed to either serum or serum-free media. The serum-free media was comprised solely of Dulbecco's Modified Eagle's (DME) High Glucose Medium (Irvine Scientific #9024). In addition to the above DME Medium, the media with serum contained the following: 10% Fetal Bovine Serum (Hyclone #SH30071.03), 1% of 100mM Sodium Pyruvate (Irvine Scientific #9334), 1% of 20mM L-Glutamine (Irvine Scientific #9317), and 1% of Penicillin-

Streptomycin solution (Irvine Scientific #9366).

min min

in in

1. L. L. L. M. M. 1.1.

10

15

20

A 96-well Adenylyl Cyclase Activation Flashplate™ was used (NEN: #SMP004A). First, 50ul of the standards for the assay were added to the plate, in duplicate, ranging from concentrations of 50pmol to zero pmol cAMP per well. The standard cAMP (NEN: #SMP004A) was reconstituted in water, and serial dilutions were made using 1xPBS (Irvine Scientific: #9240). Next, 50ul of the stimulation buffer (NEN: #SMP004A) was added to all wells. In the case of using compounds to measure activation or inactivation of cAMP, 10ul of each compound, diluted in water, was added to its respective well, in triplicate. Various final concentrations used range from 1uM up to 1mM. Adenosine 5'-triphosphate, ATP, (Research Biochemicals International: #A-141) and Adenosine 5'-diphosphate, ADP, (Sigma: #A2754) were used in the assay. Next, the 293 cells transfected with the respective cDNA (CMV or TDAG8) were harvested 24 (assay detection in serum media) or 48 hours posttransfection (assay detection comparing serum and serum-free media). The media was aspirated and the cells washed once with 1xPBS. Then 5ml of 1xPBS was added to the cells along with 3ml of cell dissociation buffer (Sigma: #C-1544). The detached cells were transferred to a centrifuge tube and centrifuged at room temperature for five minutes. The supernatant was removed and the cell pellet was resuspended in an appropriate amount of 1xPBS to obtain a final concentration of 2x10⁶ cells per milliliter. To the wells containing the compound, 50ul of the cells in 1xPBS (1x10⁵ cells/well) were added. The plate was incubated on a shaker for 15 minutes at room temperature. The detection buffer containing the tracer cAMP was prepared. In 11ml of detection buffer (NEN: #SMP004A), 50ul (equal to 1uCi) of [125] cAMP (NEN: #SMP004A) was added. Following incubation, 50ul of this detection buffer containing tracer cAMP was added to each well. The plate was placed on a shaker and Late the contract of the contr

15

incubated at room temperature for two hours. Finally, the solution from the wells of the plate were aspirated and the flashplate was counted using the Wallac MicroBeta™ scintillation counter.

In Figure 2A, ATP and ADP bind to endogenous TDAG8 resulting in an increase of cAMP of about 59% and about 55% respectively. Figure 2B evidences ATP and ADP binding to endogenous TDAG8 where endogenous TDAG8 was transfected and grown in serum and serum-free medium. ATP binding to endogenous TDAG8 grown in serum media evidences an increase in cAMP of about 65%, compared to the endogenous TDAG8 with no compounds; in serum-free media there was an increase of about 68%. ADP binding to endogenous TDAG8 in serum evidences about a 61% increase, while in serum-free ADP binding evidences an increase of about 62% increase. ATP and ADP bind to endogenous TDAG8 with an EC50 value of 139.8uM and 120.5uM, respectively (data not shown).

Although the results presented in Figure 2B indicate substantially the same results when serum and serum-free media were compared, our choice is to use a serum based media, although a serum-free media can also be utilized.

Example 6 GPCR FUSION PROTEIN PREPARATION

The design of the constitutively activated GPCR-G protein fusion construct was accomplished as follows: both the 5' and 3' ends of the rat G protein Gsα (long form; Itoh, H. et al., 83 *PNAS* 3776 (1986)) were engineered to include a HindIII (5'-AAGCTT-3') sequence thereon. Following confirmation of the correct sequence (including the flanking HindIII sequences), the entire sequence was shuttled into pcDNA3.1(-) (Invitrogen, cat. no. V795-20) by subcloning using the HindIII restriction site of that vector. The correct

10

20

orientation for the Gsα sequence was determined after subcloning into pcDNA3.1(-). The modified pcDNA3.1(-) containing the rat Gsα gene at HindIII sequence was then verified; this vector was now available as a "universal" Gsα protein vector. The pcDNA3.1(-) vector contains a variety of well-known restriction sites upstream of the HindIII site, thus beneficially providing the ability to insert, upstream of the Gs protein, the coding sequence of an endogenous, constitutively active GPCR. This same approach can be utilized to create other "universal" G protein vectors, and, of course, other commercially available or proprietary vectors known to the artisan can be utilized – the important criteria is that the sequence for the GPCR be upstream and in-frame with that of the G protein.

TDAG8 couples via Gs, while H9 couples via Gz. For the following exemplary GPCR Fusion Proteins, fusion to Gsa was accomplished.

A TDAG8(I225K)-Gsα Fusion Protein construct was made as follows: primers were designed as follows:

- 5'-gatcTCTAGAATGAACAGCACATGTATTGAAG-3' (SEQ.ID.NO.: 125; sense)
- 15 5'-ctagGGTACCCGCTCAAGGACCTCTAATTCCATAG-3' (SEQ.ID.NO.: 126; antisense).

Nucleotides in lower caps are included as spacers in the restriction sites between the G protein and TDAG8. The sense and anti-sense primers included the restriction sites for XbaI and KpnI, respectively.

PCR was then utilized to secure the respective receptor sequences for fusion within the Gsα universal vector disclosed above, using the following protocol for each: 100ng cDNA for TDAG8 was added to separate tubes containing 2ul of each primer (sense and anti-sense), 3uL of 10mM dNTPs, 10uL of 10XTaqPlus™ Precision buffer, 1uL of TaqPlus™ Precision polymerase (Stratagene: #600211), and 80uL of water. Reaction temperatures and cycle times for TDAG8 were as follows: the initial denaturing step was done it 94°C for five minutes, and

15

a cycle of 94°C for 30 seconds; 55°C for 30 seconds; 72°C for two minutes. A final extension time was done at 72°C for ten minutes. PCR product for was run on a 1% agarose gel and then purified (data not shown). The purified product was digested with XbaI and KpnI (New England Biolabs) and the desired inserts purified and ligated into the Gs universal vector at the respective restriction site. The positive clones was isolated following transformation and determined by restriction enzyme digest; expression using 293 cells was accomplished following the protocol set forth *infra*. Each positive clone for TDAG8:Gs – Fusion Protein was sequenced to verify correctness.

GPCR Fusion Proteins comprising non-endogenous, constitutively activated

TDAG8(I225K) were analyzed as above and verified for constitutive activation.

An H9(F236K)-Gsα Fusion Protein construct was made as follows: primers were designed as follows:

- 5'-TTAgatatcGGGGCCCACCCTAGCGGT-3' (SEQ.ID.NO.: 145; sense)
- 5'-ggtaccCCCACAGCCATTTCATCAGGATC-3' (SEQ.ID.NO.: 146; antisense).

Nucleotides in lower caps are included as spacers in the restriction sites between the G protein and H9. The sense and anti-sense primers included the restriction sites for EcoRV and KpnI, respectively such that spacers (attributed to the restriction sites) exists between the G protein and H9.

PCR was then utilized to secure the respective receptor sequences for fusion within the Gsα universal vector disclosed above, using the following protocol for each: 80ng cDNA for H9 was added to separate tubes containing 100ng of each primer (sense and anti-sense), and 45uL of PCR SupermixTM (Gibco-Brl, LifeTech) (50ul total reaction volume). Reaction temperatures and cycle times for H9 were as follows: the initial denaturing step was done it 94°C for one, and a cycle of 94°C for 30 seconds; 55°C for 30 seconds; 72°C for two

minutes. A final extension time was done at 72 °C for seven minutes. PCR product for was

T.

mil fin

Ш

Ho is the face of the same of

To ascertain the ability of measuring a cAMP response mediated by the Gs protein (even though H9 couples with Gz), the following cAMP membrane assay was utilized, based upon an NEN Adenyl Cyclase Activation Flahplate™ Assay kit (96 well format). "Binding Buffer" consisted of 10mM HEPES, 100mM NaCl and 10mM MgCl (ph 7.4). "Regeneration Buffer" was prepared in Binding Buffer and consisted of 20mM phosphocreatine, 20U creatine phosphokinase, 20uM GTP, 0.2mM ATP, and 0.6mM IBMX. "cAMP Standards" were prepared in Binding Buffer as follows:

| | (5,000 pmol | IP Stock /ml in 2ml H ₂ O) in ul | Added to indicted amount of Binding Buffer | Final Assay Concentration (50ul into 100ul) to achieve indicated pmol/well |
|------|-------------|---|--|--|
| 20 | Α | 250 | 1ml | 50 |
| | В | 500 of A | 500ul | 25 |
| | C | 500 of B | 500ul | 12.5 |
| | Ď | 500 of C | 750ul | 5.0 |
| | E | 500 of D | 500ul | 2.5 |
| 25 | F | 500 of E | 500ul | 1.25 |
| ک سد | Ğ | 500 of F | 750ul | 0.5 |

Frozen membranes (both pCMV as control and the non-endogenous H(-Gs Fusion Protein) were thawed (on ice at room temperature until in solution). Membranes were

m IL L

10

20

homogenized with a polytron until in suspension (2 x 15 seconds). Membrane protein concentration was determined using the Bradford Assay Protocol (see infra). Membrane concentration was diluted to 0.5mg/ml in Regeneration Buffer (final assay concentration -25ug/well). Thereafter, 50ul of Binding Buffer was added to each well. For control, 50ul/well of cAMP standard was added to wells 11 and 12 A-G, with Binding Buffer alone to 12H (on the 96-well format). Thereafter, 50ul/well of protein was added to the wells and incubated at room temperature (on shaker) for 60min. 100ul[125I]cAMP in Detection Buffer (see infra) was added to each well (final - 50ul[125]]cAMP into 11ml Detection Buffer). These were incubated for 2hrs at room temperature. Plates were aspirated with an 8 channel manifold and sealed with plate covers. Results (pmoles cAMP bound) were read in a Wallac™ 1450 on "prot #15). Results are presented in Figure 3.

The results presented in Figure 3 indicate that the Gs coupled fusion was able to "drive" the cyclase reaction such that measurement of the consitutive activation of H9(F236K) was viable. Based upon these results, the direct identification of candidate compounds that are inverse agonists, agonists and partial agonists is possible using a cyclase-based assay.

Example 6

Protocol: Direct Identification of Inverse Agonists and Agonists Using [35S]GTPγS

Although we have utilized endogenous, constitutively active GPCRs for the direct identification of candidate compounds as, e.g., inverse agonists, for reasons that are not altogether understood, intra-assay variation can become exacerbated. Preferably, then, a GPCR Fusion Protein, as disclosed above, is also utilized with a non-endogenous, constitutively activated GPCR. We have determined that when such a protein is used, intraassay variation appears to be substantially stabilized, whereby an effective signal-to-noise ratio is obtained. This has the beneficial result of allowing for a more robust identification of candidate compounds. Thus, it is preferred that for direct identification, a GPCR Fusion Protein be used and that when utilized, the following assay protocols be utilized.

Membrane Preparation

Membranes comprising the non-endogenous, constitutively active orphan GPCR

Fusion Protein of interest and for use in the direct identification of candidate compounds as inverse agonists, agonists or partial agonists are preferably prepared as follows:

a. Materials

50.74

10

15

20

 "Membrane Scrape Buffer" is comprised of 20mM HEPES and 10mM EDTA, pH 7.4;

"Membrane Wash Buffer" is comprised of 20 mM HEPES and 0.1 mM EDTA, pH 7.4;

"Binding Buffer" is comprised of 20mM HEPES, 100 mM NaCl, and 10 mM MgCl₂, pH 7.4

b. Procedure

All materials are kept on ice throughout the procedure. Firstly, the media is aspirated from a confluent monolayer of cells, followed by rinse with 10ml cold PBS, followed by aspiration. Thereafter, 5ml of Membrane Scrape Buffer is added to scrape cells; this is followed by transfer of cellular extract into 50ml centrifuge tubes (centrifuged at 20,000 rpm for 17 minutes at 4°C). Thereafter, the supernatant is aspirated and the pellet is resuspended in 30ml Membrane Wash Buffer followed by centrifuge at 20,000 rpm for 17 minutes at 4°C. The supernatant is then aspirated and the pellet resuspended in Binding Buffer. This is then homogenized using a Brinkman polytronTM homogenizer (15-20 second bursts until the all material is in suspension). This is referred to herein as "Membrane Protein".

Bradford Protein Assay

Following the homogenization, protein concentration of the membranes is determined using the Bradford Protein Assay (protein can be diluted to about 1.5mg/ml, aliquoted and

10

20

frozen (-80°C) for later use; when frozen, protocol for use is as follows: on the day of the assay, frozen Membrane Protein is thawed at room temperature, followed by vortex and then homogenized with a polytron at about 12 x 1,000 rpm for about 5-10 seconds; it is noted that for multiple preparations, the homogenizor should be thoroughly cleaned between homogenezation of different preparations).

a. Materials

Binding Buffer (as per above); Bradford Dye Reagent; Bradford Protein Standard are utilized, following manufacturer instructions (Biorad, cat. no. 500-0006).

b. Procedure

Duplicate tubes are prepared, one including the membrane, and one as a control "blank". Each contained 800ul Binding Buffer. Thereafter, 10ul of Bradford Protein Standard (1mg/ml) is added to each tube, and 10ul of membrane Protein is then added to just one tube (not the blank). Thereafter, 200ul of Bradford Dye Reagent is added to each tube, followed by vortex of each. After five (5) minutes, the tubes were re-vortexed and the material therein is transferred to cuvettes. The cuvettes are then read using a CECIL 3041 spectrophotometer, at wavelength 595.

Direct Identification Assay

a. Materials

GDP Buffer consists of 37.5 ml Binding Buffer and 2mg GDP (Sigma, cat. no. G-7127), followed by a series of dilutions in Binding Buffer to obtain 0.2 uM GDP (final concentration of GDP in each well was 0.1 uM GDP); each well comprising a candidate compound, has a final volume of 200ul consisting of 100ul GDP Buffer (final concentration, 0.1uM GDP), 50ul Membrane Protein in Binding Buffer, and 50ul [35S]GTPγS (0.6 nM) in

Binding Buffer (2.5 ul [35S]GTPγS per 10ml Binding Buffer).

Procedure b.

Candidate compounds are preferably screened using a 96-well plate format (these can be frozen at -80°C). Membrane Protein (or membranes with expression vector excluding the 5 GPCR Fusion Protein, as control), are homogenized briefly until in suspension. Protein concentration is then determined using the Bradford Protein Assay set forth above. Membrane Protein (and control) is then diluted to 0.25mg/ml in Binding Buffer (final assay concentration, 12.5ug/well). Thereafter, 100 ul GDP Buffer is added to each well of a Wallac Scintistrip™ (Wallac). A 5ul pin-tool is then used to transfer 5 ul of a candidate compound into such well (i.e., 5ul in total assay volume of 200 ul is a 1:40 ratio such that the final screening concentration of the candidate compound is 10uM). Again, to avoid contamination, after each transfer step the pin tool should be rinsed in three reservoirs comprising water (1X), ethanol (1X) and water (2X) - excess liquid should be shaken from the tool after each rinse and dried with paper and kimwipes. Thereafter, 50 ul of Membrane Protein is added to each well (a control well comprising membranes without the GPCR Fusion Protein is also utilized), and pre-incubated for 5-10 minutes at room temperature. Thereafter, 50 ul of [35S]GTPγS (0.6 nM) in Binding Buffer is added to each well, followed by incubation on a shaker for 60 minutes at room temperature (again, in this example, plates were covered with foil). The assay is then stopped by spinning of the plates at 4000 RPM for 15 minutes at 22°C. The plates are then aspirated with an 8 channel manifold and sealed with plate covers. The plates are then read on a Wallacc 1450 using setting "Prot. #37" (as per manufacturer instructions).

Example 7

20

T H Hay mal

A Tall bull dan

Protocol: Confirmation Assay

Using an independent assay approach to provide confirmation of a directly identified

candidate compound as set forth above, it is preferred that a confirmation assay then be utilized. In this case, the preferred confirmation assay is a cyclase-based assay.

A modified Flash Plate™ Adenylyl Cyclase kit (New England Nuclear; Cat. No. SMP004A) is preferably utilized for confirmation of candidate compounds directly identified as inverse agonists and agonists to non-endogenous, constitutively activated orphan GPCRs in accordance with the following protocol.

Transfected cells are harvested approximately three days after transfection. Membranes are prepared by homogenization of suspended cells in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCl₂. Homogenization is performed on ice using a Brinkman Polytron™ for approximately 10 seconds. The resulting homogenate is centrifuged at 49,000 X g for 15 minutes at 4°C. The resulting pellet is then resuspended in buffer containing 20mM HEPES, pH 7.4 and 0.1 mM EDTA, homogenized for 10 seconds, followed by centrifugation at 49,000 X g for 15 minutes at 4°C. The resulting pellet can be stored at -80°C until utilized. On the day of direct identification screening, the membrane pellet is slowly thawed at room temperature, resuspended in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCL2, to yield a final protein concentration of 0.60mg/ml (the resuspended membranes are placed on ice until use).

the last the last the real last the las

10

15

20

cAMP standards and Detection Buffer (comprising 2 μ Ci of tracer [125 I cAMP (100 μ l] to 11 ml Detection Buffer) are prepared and maintained in accordance with the manufacturer's instructions. Assay Buffer is prepared fresh for screening and contained 20mM HEPES, pH 7.4, 10mM MgCl₂, 20mM phospocreatine (Sigma), 0.1 units/ml creatine phosphokinase (Sigma), 50 μ M GTP (Sigma), and 0.2 mM ATP (Sigma); Assay Buffer can be stored on ice until utilized.

In the last of the same one in the same one one or the same of the

5

10

20

Candidate compounds identified as per above (if frozen, thawed at room temperature) are added, preferably, to 96-well plate wells (3μ l/well; 12μ M final assay concentration), together with 40μ l Membrane Protein (30μ g/well) and 50μ l of Assay Buffer. This admixture is then incubated for 30 minutes at room temperature, with gentle shaking.

Following the incubation, 100µl of Detection Buffer is added to each well, followed by incubation for 2-24 hours. Plates are then counted in a Wallac MicroBeta™ plate reader using "Prot. #31" (as per manufacturer instructions).

It is intended that each of the patents, applications, and printed publications mentioned in this patent document be hereby incorporated by reference in their entirety.

As those skilled in the art will appreciate, numerous changes and modifications may be made to the preferred embodiments of the invention without departing from the spirit of the invention. It is intended that all such variations fall within the scope of the invention.

Although a variety of expression vectors are available to those in the art, for purposes of utilization for both the endogenous and non-endogenous human GPCRs, it is most preferred that the vector utilized be pCMV. This vector was deposited with the American Type Culture Collection (ATCC) on October 13, 1998 (10801 University Blvd., Manassas, VA 20110-2209 USA) under the provisions of the Budapest Treaty for the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure. The DNA was tested by the ATCC and determined to be. The ATCC has assigned the following deposit number to pCMV: ATCC #203351.

SEQUENCE LISTING

10

20

25

30

(1) GENERAL INFORMATION:

(i) APPLICANT: Behan, Dominic P.

Lehmann-Bruinsma, Karin

5 Chalmers, Derek T.

Lowitz, Kevin P.

Lin, I-Lin

Dang, Huong T.

Chen, Ruoping

Liaw, Chen W.

Gore, Martin J.

White, Carol

(ii) TITLE OF INVENTION: Non-Endogenous, Constitutively Activated Human G Protein-Coupled Receptors

15 (iii) NUMBER OF SEQUENCES: 146

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Arena Pharmaceuticals, Inc.
- (B) STREET: 6166 Nancy Ridge Drive
- (C) CITY: San Diego
- (D) STATE: CA
- (E) COUNTRY: USA
- (F) ZIP: 92121

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US

- (B) FILING DATE:
- (C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Burgoon, Richard P.
- 35 (B) REGISTRATION NUMBER: 34,787

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (858)453-7200
- (B) TELEFAX: (858) 453-7210

(2) INFORMATION FOR SEQ ID NO:1:

40 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1260 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

| | ATGGTCTTCT | CGGCAGTGTT | GACTGCGTTC | CATACCGGGA | CATCCAACAC | AACATTTGTC | 60 |
|----|------------|------------|------------|------------|------------|------------|-----|
| 5 | GTGTATGAAA | ACACCTACAT | GAATATTACA | CTCCCTCCAC | CATTCCAGCA | TCCTGACCTC | 120 |
| | AGTCCATTGC | TTAGATATAG | TTTTGAAACC | ATGGCTCCCA | CTGGTTTGAG | TTCCTTGACC | 180 |
| | GTGAATAGTA | CAGCTGTGCC | CACAACACCA | GCAGCATTTA | AGAGCCTAAA | CTTGCCTCTT | 240 |
| | CAGATCACCC | TTTCTGCTAT | AATGATATTC | ATTCTGTTTG | TGTCTTTTCT | TGGGAACTTG | 300 |
| | GTTGTTTGCC | TCATGGTTTA | CCAAAAAGCT | GCCATGAGGT | CTGCAATTAA | CATCCTCCTT | 360 |
| 10 | GCCAGCCTAG | CTTTTGCAGA | CATGTTGCTT | GCAGTGCTGA | ACATGCCCTT | TGCCCTGGTA | 420 |
| | ACTATTCTTA | CTACCCGATG | GATTTTTGGG | AAATTCTTCT | GTAGGGTATC | TGCTATGTTT | 480 |
| | TTCTGGTTAT | TTGTGATAGA | AGGAGTAGCC | ATCCTGCTCA | TCATTAGCAT | AGATAGGTTC | 540 |
| | CTTATTATAG | TCCAGAGGCA | GGATAAGCTA | AACCCATATA | GAGCTAAGGT | TCTGATTGCA | 600 |
| | GTTTCTTGGG | CAACTTCCTT | TTGTGTAGCT | TTTCCTTTAG | CCGTAGGAAA | CCCCGACCTG | 660 |
| 15 | CAGATACCTT | CCCGAGCTCC | CCAGTGTGTG | TTTGGGTACA | CAACCAATCC | AGGCTACCAG | 720 |
| | GCTTATGTGA | TTTTGATTTC | TCTCATTTCT | TTCTTCATAC | CCTTCCTGGT | AATACTGTAC | 780 |
| | TCATTTATGG | GCATACTCAA | CACCCTTCGG | CACAATGCCT | TGAGGATCCA | TAGCTACCCT | 840 |
| | GAAGGTATAT | GCCTCAGCCA | GGCCAGCAAA | CTGGGTCTCA | TGAGTCTGCA | GAGACCTTTC | 900 |
| | CAGATGAGCA | TTGACATGGG | CTTTAAAACA | CGTGCCTTCA | CCACTATTT | GATTCTCTTT | 960 |
| 20 | GCTGTCTTCA | TTGTCTGCTG | GGCCCCATTC | ACCACTTACA | GCCTTGTGGC | AACATTCAGT | 102 |
| | AAGCACTTTT | ACTATCAGCA | CAACTTTTTT | GAGATTAGCA | CCTGGCTACT | GTGGCTCTGC | 108 |
| | TACCTCAAGT | CTGCATTGAA | TCCGCTGATC | TACTACTGGA | GGATTAAGAA | ATTCCATGAT | 114 |
| | GCTTGCCTGG | ACATGATGCC | TAAGTCCTTC | AAGTTTTTGC | CGCAGCTCCC | TGGTCACACA | 120 |
| | AAGCGACGGA | тасстсстас | TGCTGTCTAT | GTGTGTGGGG | AACATCGGAC | GGTGGTGTGA | 126 |

25 (3) INFORMATION FOR SEQ ID NO:2:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 419 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

| Met | Val | Phe | Ser | Ala | Val | Leu | Thr | Ala | Phe | His | Thr | Gly | Thr | Ser | Asn |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |

- 5 Thr Thr Phe Val Val Tyr Glu Asn Thr Tyr Met Asn Ile Thr Leu Pro
 - Pro Pro Phe Gln His Pro Asp Leu Ser Pro Leu Leu Arg Tyr Ser Phe 35 40 45
- Glu Thr Met Ala Pro Thr Gly Leu Ser Ser Leu Thr Val Asn Ser Thr 50 55 60
 - Ala Val Pro Thr Thr Pro Ala Ala Phe Lys Ser Leu Asn Leu Pro Leu 65 70 75 80
 - Gln Ile Thr Leu Ser Ala Ile Met Ile Phe Ile Leu Phe Val Ser Phe 85 90 95
- Leu Gly Asn Leu Val Val Cys Leu Met Val Tyr Gln Lys Ala Ala Met
 100 105 110
 - Arg Ser Ala Ile Asn Ile Leu Leu Ala Ser Leu Ala Phe Ala Asp Met 115 120 125
- Leu Leu Ala Val Leu Asn Met Pro Phe Ala Leu Val Thr Ile Leu Thr 20 130 135 140
 - Thr Arg Trp Ile Phe Gly Lys Phe Phe Cys Arg Val Ser Ala Met Phe 145 150 155 160
 - Phe Trp Leu Phe Val Ile Glu Gly Val Ala Ile Leu Leu Ile Ile Ser
- 25 Ile Asp Arg Phe Leu Ile Ile Val Gln Arg Gln Asp Lys Leu Asn Pro 180 185 190
 - Tyr Arg Ala Lys Val Leu Ile Ala Val Ser Trp Ala Thr Ser Phe Cys 195 200 205
- Val Ala Phe Pro Leu Ala Val Gly Asn Pro Asp Leu Gln Ile Pro Ser 30 210 215 220
 - Arg Ala Pro Gln Cys Val Phe Gly Tyr Thr Thr Asn Pro Gly Tyr Gln 225 230 235 240
 - Ala Tyr Val Ile Leu Ile Ser Leu Ile Ser Phe Phe Ile Pro Phe Leu 245 250 255
- Val Ile Leu Tyr Ser Phe Met Gly Ile Leu Asn Thr Leu Arg His Asn 260 265 270

| 41017 | |
|---------|---------------------|
| Ť | Ī |
| mm. | Ţ. |
| ÷ | 5 <u>.</u> 2 |
| .111111 | Fi |
| 1211112 | U |
| .4444. | F |
| 1121211 | L. |
| ë | |
| .41114. | ## P |
| ä | Ţ |
| THE P | 027 <u>1</u> 227 |
| | 14 |
| anni. | 725 725 |
| mini | nŠt |

Ala Leu Arg Ile His Ser Tyr Pro Glu Gly Ile Cys Leu Ser Gln Ala 275 280 Ser Lys Leu Gly Leu Met Ser Leu Gln Arg Pro Phe Gln Met Ser Ile 295 Asp Met Gly Phe Lys Thr Arg Ala Phe Thr Thr Ile Leu Ile Leu Phe 5 305 Ala Val Phe Ile Val Cys Trp Ala Pro Phe Thr Thr Tyr Ser Leu Val 325 330 Ala Thr Phe Ser Lys His Phe Tyr Tyr Gln His Asn Phe Phe Glu Ile 10 Ser Thr Trp Leu Leu Trp Leu Cys Tyr Leu Lys Ser Ala Leu Asn Pro 360 Leu Ile Tyr Tyr Trp Arg Ile Lys Lys Phe His Asp Ala Cys Leu Asp 375 Met Met Pro Lys Ser Phe Lys Phe Leu Pro Gln Leu Pro Gly His Thr 15 385 Lys Arg Arg Ile Arg Pro Ser Ala Val Tyr Val Cys Gly Glu His Arg 410 405 Thr Val Val 20 (4) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1119 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

| | ATGTTAGCCA | ACAGCTCCTC | AACCAACAGT | TCTGTTCTCC | CGTGTCCTGA | CTACCGACCT | 60 |
|----|------------|------------|------------|------------|------------|------------|-----|
| 30 | ACCCACCGCC | TGCACTTGGT | GGTCTACAGC | TTGGTGCTGG | CTGCCGGGCT | CCCCTCAAC | 120 |
| | GCGCTAGCCC | TCTGGGTCTT | CCTGCGCGCG | CTGCGCGTGC | ACTCGGTGGT | GAGCGTGTAC | 180 |
| | ATGTGTAACC | TGGCGGCCAG | CGACCTGCTC | TTCACCCTCT | CGCTGCCCGT | TCGTCTCTCC | 240 |
| | TACTACGCAC | TGCACCACTG | GCCCTTCCCC | GACCTCCTGT | GCCAGACGAC | GGGCGCCATC | 300 |
| | TTCCAGATGA | ACATGTACGG | CAGCTGCATC | TTCCTGATGC | TCATCAACGT | GGACCGCTAC | 360 |

| | GCCGCCATCG | TGCACCCGCT | GCGACTGCGC | CACCTGCGGC | GGCCCCGCGT | GGCGCGGCTG | 420 |
|----|------------|------------|------------|------------|------------|------------|------|
| | CTCTGCCTGG | GCGTGTGGGC | GCTCATCCTG | GTGTTTGCCG | TGCCCGCCGC | CCGCGTGCAC | 480 |
| | AGGCCCTCGC | GTTGCCGCTA | CCGGGACCTC | GAGGTGCGCC | TATGCTTCGA | GAGCTTCAGC | 540 |
| | GACGAGCTGT | GGAAAGGCAG | GCTGCTGCCC | CTCGTGCTGC | TGGCCGAGGC | GCTGGGCTTC | 600 |
| 5 | CTGCTGCCCC | TGGCGGCGGT | GGTCTACTCG | TCGGGCCGAG | TCTTCTGGAC | GCTGGCGCGC | 660 |
| | CCCGACGCCA | CGCAGAGCCA | GCGGCGGCGG | AAGACCGTGC | GCCTCCTGCT | GGCTAACCTC | 720 |
| | GTCATCTTCC | TGCTGTGCTT | CGTGCCCTAC | AACAGCACGC | TGGCGGTCTA | CGGGCTGCTG | 780 |
| | CGGAGCAAGC | TGGTGGCGGC | CAGCGTGCCT | GCCCGCGATC | GCGTGCGCGG | GGTGCTGATG | 840 |
| | GTGATGGTGC | TGCTGGCCGG | CGCCAACTGC | GTGCTGGACC | CGCTGGTGTA | CTACTTTAGC | 900 |
| 10 | GCCGAGGGCT | TCCGCAACAC | CCTGCGCGGC | CTGGGCACTC | CGCACCGGGC | CAGGACCTCG | 960 |
| | GCCACCAACG | GGACGCGGGC | GGCGCTCGCG | CAATCCGAAA | GGTCCGCCGT | CACCACCGAC | 1020 |
| | GCCACCAGGC | CGGATGCCGC | CAGTCAGGGG | CTGCTCCGAC | CCTCCGACTC | CCACTCTCTG | 1080 |
| | TCTTCCTTCA | CACAGTGTCC | CCAGGATTCC | GCCCTCTGA | | | 1119 |

- (5) INFORMATION FOR SEQ ID NO:4:
- 15 (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 372 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- 20 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Leu Ala Asn Ser Ser Ser Thr Asn Ser Ser Val Leu Pro Cys Pro 1 5 10 15

Asp Tyr Arg Pro Thr His Arg Leu His Leu Val Val Tyr Ser Leu Val 25 20 25 30

Leu Ala Ala Gly Leu Pro Leu Asn Ala Leu Ala Leu Trp Val Phe Leu 35 40 45

Arg Ala Leu Arg Val His Ser Val Val Ser Val Tyr Met Cys Asn Leu 50 55 60

30 Ala Ala Ser Asp Leu Leu Phe Thr Leu Ser Leu Pro Val Arg Leu Ser 65 70 75 80

Tyr Tyr Ala Leu His His Trp Pro Phe Pro Asp Leu Leu Cys Gln Thr

85 90 95 Thr Gly Ala Ile Phe Gln Met Asn Met Tyr Gly Ser Cys Ile Phe Leu 105 100 Met Leu Ile Asn Val Asp Arg Tyr Ala Ala Ile Val His Pro Leu Arg 5 120 Leu Arg His Leu Arg Arg Pro Arg Val Ala Arg Leu Leu Cys Leu Gly 130 135 Val Trp Ala Leu Ile Leu Val Phe Ala Val Pro Ala Ala Arg Val His Arg Pro Ser Arg Cys Arg Tyr Arg Asp Leu Glu Val Arg Leu Cys Phe 10 170 Glu Ser Phe Ser Asp Glu Leu Trp Lys Gly Arg Leu Leu Pro Leu Val 180 185 Leu Leu Ala Glu Ala Leu Gly Phe Leu Leu Pro Leu Ala Ala Val Val 15 200 Tyr Ser Ser Gly Arg Val Phe Trp Thr Leu Ala Arg Pro Asp Ala Thr 215 210 Gln Ser Gln Arg Arg Arg Lys Thr Val Arg Leu Leu Leu Ala Asn Leu Val Ile Phe Leu Leu Cys Phe Val Pro Tyr Asn Ser Thr Leu Ala Val 20 250 Tyr Gly Leu Leu Arg Ser Lys Leu Val Ala Ala Ser Val Pro Ala Arg 265 260 Asp Arg Val Arg Gly Val Leu Met Val Met Val Leu Leu Ala Gly Ala 280 25 Asn Cys Val Leu Asp Pro Leu Val Tyr Tyr Phe Ser Ala Glu Gly Phe 290 Arg Asn Thr Leu Arg Gly Leu Gly Thr Pro His Arg Ala Arg Thr Ser 310 Ala Thr Asn Gly Thr Arg Ala Ala Leu Ala Gln Ser Glu Arg Ser Ala 30 330 Val Thr Thr Asp Ala Thr Arg Pro Asp Ala Ala Ser Gln Gly Leu Leu 345 340 Arg Pro Ser Asp Ser His Ser Leu Ser Ser Phe Thr Gln Cys Pro Gln 365 360 35 Asp Ser Ala Leu

370

(6) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1107 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

| | ATGGCCAACT | CCACAGGGCT | GAACGCCTCA | GAAGTCGCAG | GCTCGTTGGG | GTTGATCCTG | 60 |
|----|------------|--------------|------------|------------|------------|------------|-------|
| 10 | GCAGCTGTCG | TGGAGGTGGG | GGCACTGCTG | GGCAACGGCG | CGCTGCTGGT | CGTGGTGCTG | 120 |
| | CGCACGCCGG | GACTGCGCGA | CGCGCTCTAC | CTGGCGCACC | TGTGCGTCGT | GGACCTGCTG | 180 |
| | GCGGCCGCCT | CCATCATGCC | GCTGGGCCTG | CTGGCCGCAC | cgccgcccgg | GCTGGGCCGC | 240 |
| | GTGCGCCTGG | GCCCGCGCC | ATGCCGCGCC | GCTCGCTTCC | TCTCCGCCGC | TCTGCTGCCG | ; 300 |
| | GCCTGCACGC | TCGGGGTGGC | CGCACTTGGC | CTGGCACGCT | ACCGCCTCAT | CGTGCACCCG | 360 |
| 15 | CTGCGGCCAG | GCTCGCGGCC | GCCGCCTGTG | CTCGTGCTCA | CCGCCGTGTG | GGCCGCGGCG | 420 |
| | GGACTGCTGG | GCGCGCTCTC | CCTGCTCGGC | CCGCCGCCCG | CACCGCCCCC | TGCTCCTGCT | 480 |
| | CGCTGCTCGG | TCCTGGCTGG | GGGCCTCGGG | CCCTTCCGGC | CGCTCTGGGC | CCTGCTGGCC | 540 |
| | TTCGCGCTGC | CCGCCCTCCT | GCTGCTCGGC | GCCTACGGCG | GCATCTTCGT | GGTGGCGCGT | 600 |
| | CGCGCTGCCC | TGAGGCCCCC | ACGGCCGGCG | CGCGGGTCCC | GACTCCGCTC | GGACTCTCTG | 660 |
| 20 | GATAGCCGCC | TTTCCATCTT | GCCGCCGCTC | CGGCCTCGCC | TGCCCGGGGG | CAAGGCGGCC | 720 |
| | CTGGCCCCAG | CGCTGGCCGT | GGGCCAATTT | GCAGCCTGCT | GGCTGCCTTA | TGGCTGCGCG | 780 |
| | TGCCTGGCGC | CCGCAGCGCG | GGCCGCGGAA | GCCGAAGCGG | CTGTCACCTG | GGTCGCCTAC | 840 |
| | TCGGCCTTCG | CGGCTCACCC | CTTCCTGTAC | GGGCTGCTGC | AGCGCCCCGT | GCGCTTGGCA | 900 |
| | CTGGGCCGCC | TCTCTCGCCG | TGCACTGCCT | GGACCTGTGC | GGGCCTGCAC | TCCGCAAGCC | 960 |
| 25 | TGGCACCCGC | : GGGCACTCTT | GCAATGCCTC | CAGAGACCCC | CAGAGGGCCC | TGCCGTAGGC | 1020 |
| | CCTTCTGAGG | CTCCAGAACA | GACCCCCGAG | TTGGCAGGAG | GGCGGAGCCC | CGCATACCAG | 1080 |
| | GGGCCACCTG | G AGAGTTCTCT | CTCCTGA | | | | 1107 |

(7) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 368 amino acids 30

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Ala Asn Ser Thr Gly Leu Asn Ala Ser Glu Val Ala Gly Ser Leu

Gly Leu Ile Leu Ala Ala Val Val Glu Val Gly Ala Leu Leu Gly Asn 20 25 30

10 Gly Ala Leu Leu Val Val Leu Arg Thr Pro Gly Leu Arg Asp Ala 35 40 45

Leu Tyr Leu Ala His Leu Cys Val Val Asp Leu Leu Ala Ala Ala Ser 50 55 60

Ile Met Pro Leu Gly Leu Leu Ala Ala Pro Pro Pro Gly Leu Gly Arg
65 70 75 80

Val Arg Leu Gly Pro Ala Pro Cys Arg Ala Ala Arg Phe Leu Ser Ala 85 90 95

Ala Leu Leu Pro Ala Cys Thr Leu Gly Val Ala Ala Leu Gly Leu Ala
100 105 110

20 Arg Tyr Arg Leu Ile Val His Pro Leu Arg Pro Gly Ser Arg Pro Pro 115 120 125

Pro Val Leu Val Leu Thr Ala Val Trp Ala Ala Ala Gly Leu Leu Gly 130 135 140

Ala Leu Ser Leu Leu Gly Pro Pro Pro Ala Pro Pro Pro Ala Pro Ala 25 150 155 160

Arg Cys Ser Val Leu Ala Gly Gly Leu Gly Pro Phe Arg Pro Leu Trp 165 170 175

Ala Leu Leu Ala Phe Ala Leu Pro Ala Leu Leu Leu Leu Gly Ala Tyr 180 185 190

30 Gly Gly Ile Phe Val Val Ala Arg Arg Ala Ala Leu Arg Pro Pro Arg 195 200 205

Pro Ala Arg Gly Ser Arg Leu Arg Ser Asp Ser Leu Asp Ser Arg Leu 210 215 220

Ser Ile Leu Pro Pro Leu Arg Pro Arg Leu Pro Gly Gly Lys Ala Ala 35 225 230 235 240

Leu Ala Pro Ala Leu Ala Val Gly Gln Phe Ala Ala Cys Trp Leu Pro

| i Di |
|------|
| 罰 |
| 1 |
| ij. |
| |
| Ţ |
| Per |
| £ |
| 1 |
| m |
| |
| 1 |
| |
| |

| | | | | | | 245 | | | | | 250 | | | | | 255 | |
|----|-----|------------|----------------|----------------------|---------------------|--------------|--------------------|---------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | | | | | | 243 | | | | | 200 | | | | | | |
| | | Tyr | Gly | Cys | Ala 260 | Cys | Leu | Ala | Pro | Ala 265 | Ala | Arg | Ala | Ala | Glu 270 | Ala | Glu |
| 5 | | Ala | Ala | Val 275 | Thr | Trp | Val | Ala | Tyr 280 | Ser | Ala | Phe | Ala | Ala 285 | His | Pro | Phe |
| | | Leu | Tyr 290 | Gly | Leu | Leu | Gln | Arg 295 | Pro | Val | Arg | Leu | Ala 300 | Leu | Gly | Arg | Leu |
| | | Ser 305 | Arg | Arg | Ala | Leu | Pro 310 | Gly | Pro | Val | Arg | Ala 315 | Cys | Thr | Pro | Gln | Ala 320 |
| 10 | | Trp | His | Pro | Arg | Ala 325 | Leu | Leu | Gln | Cys | Leu 330 | Gln | Arg | Pro | Pro | Glu 335 | Gly |
| | | Pro | Ala | Val | Gly 340 | Pro | Ser | Glu | Ala | Pro 345 | Glu | Gln | Thr | Pro | Glu 350 | Leu | Ala |
| 15 | | Gly | Gly | Arg 355 | Ser | Pro | Ala | Tyr | Gln 360 | Gly | Pro | Pro | Glu | Ser 365 | Ser | Leu | Ser |
| | (8) | INFO | RMAT | ION | FOR | SEQ | ID N | 0:7: | | | | | | | | | |
| 20 | | (i) | (A (B (C | UENC) LE) TY !) ST | NGTH PE: RAND | : 10 nucl | 08 b eic SS: | ase acid sing | pair | s | | | | | | | |
| | | (ii) | MOL | ECUL | E TY | PE: | DNA | (gen | omic | :) | | | | | | | |

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
- ATGGAATCAT CTTTCTCATT TGGAGTGATC CTTGCTGTCC TGGCCTCCCT CATCATTGCT 60 25 ACTAACACAC TAGTGGCTGT GGCTGTGCTG CTGTTGATCC ACAAGAATGA TGGTGTCAGT 120 CTCTGCTTCA CCTTGAATCT GGCTGTGGCT GACACCTTGA TTGGTGTGGC CATCTCTGGC 180 CTACTCACAG ACCAGCTCTC CAGCCCTTCT CGGCCCACAC AGAAGACCCT GTGCAGCCTG 240 CGGATGGCAT TTGTCACTTC CTCCGCAGCT GCCTCTGTCC TCACGGTCAT GCTGATCACC 300 TTTGACAGGT ACCTTGCCAT CAAGCAGCCC TTCCGCTACT TGAAGATCAT GAGTGGGTTC 360 30 GTGGCCGGGG CCTGCATTGC CGGGCTGTGG TTAGTGTCTT ACCTCATTGG CTTCCTCCCA 420 CTCGGAATCC CCATGTTCCA GCAGACTGCC TACAAAGGGC AGTGCAGCTT CTTTGCTGTA 480 TTTCACCCTC ACTTCGTGCT GACCCTCTCC TGCGTTGGCT TCTTCCCAGC CATGCTCCTC 540 TTTGTCTTCT TCTACTGCGA CATGCTCAAG ATTGCCTCCA TGCACAGCCA GCAGATTCGA 600

| | AAGATGGAAC | ATGCAGGAGC | CATGGCTGGA | GGTTATCGAT | CCCCACGGAC | TCCCAGCGAC | 660 |
|---|------------|------------|------------|------------|------------|------------|------|
| | TTCAAAGCTC | TCCGTACTGT | GTCTGTTCTC | ATTGGGAGCT | TTGCTCTATC | CTGGACCCCC | 720 |
| | TTCCTTATCA | CTGGCATTGT | GCAGGTGGCC | TGCCAGGAGT | GTCACCTCTA | CCTAGTGCTG | 780 |
| | GAACGGTACC | TGTGGCTGCT | CGGCGTGGGC | AACTCCCTGC | TCAACCCACT | CATCTATGCC | 840 |
| 5 | TATTGGCAGA | AGGAGGTGCG | ACTGCAGCTC | TACCACATGG | CCCTAGGAGT | GAAGAAGGTG | 900 |
| | CTCACCTCAT | TCCTCCTCTT | TCTCTCGGCC | AGGAATTGTG | GCCCAGAGAG | GCCCAGGGAA | 960 |
| | AGTTCCTGTC | ACATCGTCAC | TATCTCCAGC | TCAGAGTTTG | ATGGCTAA | | 1008 |

- (9) INFORMATION FOR SEQ ID NO:8:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 335 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
- 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Glu Ser Ser Phe Ser Phe Gly Val Ile Leu Ala Val Leu Ala Ser

1 10 15

Leu Ile Ile Ala Thr Asn Thr Leu Val Ala Val Ala Val Leu Leu Leu 20 25 30

20 Ile His Lys Asn Asp Gly Val Ser Leu Cys Phe Thr Leu Asn Leu Ala 35 40 45

Val Ala Asp Thr Leu Ile Gly Val Ala Ile Ser Gly Leu Leu Thr Asp 50 55 60

Gln Leu Ser Ser Pro Ser Arg Pro Thr Gln Lys Thr Leu Cys Ser Leu 25 65 70 75 80

Arg Met Ala Phe Val Thr Ser Ser Ala Ala Ala Ser Val Leu Thr Val 85 90 95

Met Leu Ile Thr Phe Asp Arg Tyr Leu Ala Ile Lys Gln Pro Phe Arg 100 105 110

Tyr Leu Lys Ile Met Ser Gly Phe Val Ala Gly Ala Cys Ile Ala Gly
115 120 125

Leu Trp Leu Val Ser Tyr Leu Ile Gly Phe Leu Pro Leu Gly Ile Pro 130 135 140

Met Phe Gln Gln Thr Ala Tyr Lys Gly Gln Cys Ser Phe Phe Ala Val

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATGGACACTA CCATGGAAGC TGACCTGGGT GCCACTGGCC ACAGGCCCCG CACAGAGCTT

GATGATGAGG ACTCCTACCC CCAAGGTGGC TGGGACACGG TCTTCCTGGT GGCCCTGCTG

CTCCTTGGGC TGCCAGCCAA TGGGTTGATG GCGTGGCTGG CCGGCTCCCA GGCCCGGCAT

35 GGAGCTGGCA CGCGTCTGGC GCTGCTCCTG CTCAGCCTGG CCCTCTCTGA CTTCTTGTTC

- 72 -

Phe His Pro His Phe Val Leu Thr Leu Ser Cys Val Gly Phe Phe Pro

Ala Met Leu Leu Phe Val Phe Phe Tyr Cys Asp Met Leu Lys Ile Ala

150

155

170

AREN-0054

30

145

PATENT

160

60

120

180

10

15

20

25

| CTGGCAGCAG | CGGCCTTCCA | GATCCTAGAG | ATCCGGCATG | GGGGACACTG | GCCGCTGGGG | 300 |
|------------|------------|------------|------------|------------|------------|------|
| ACAGCTGCCT | GCCGCTTCTA | CTACTTCCTA | TGGGGCGTGT | CCTACTCCTC | CGGCCTCTTC | 360 |
| CTGCTGGCCG | CCCTCAGCCT | CGACCGCTGC | CTGCTGGCGC | TGTGCCCACA | CTGGTACCCT | 420 |
| GGGCACCGCC | CAGTCCGCCT | GCCCCTCTGG | GTCTGCGCCG | GTGTCTGGGT | GCTGGCCACA | 480 |
| CTCTTCAGCG | TGCCCTGGCT | GGTCTTCCCC | GAGGCTGCCG | TCTGGTGGTA | CGACCTGGTC | 540 |
| ATCTGCCTGG | ACTTCTGGGA | CAGCGAGGAG | CTGTCGCTGA | GGATGCTGGA | GGTCCTGGGG | 600 |
| GGCTTCCTGC | CTTTCCTCCT | GCTGCTCGTC | TGCCACGTGC | TCACCCAGGC | CACAGCCTGT | 660 |
| CGCACCTGCC | ACCGCCAACA | GCAGCCCGCA | GCCTGCCGGG | GCTTCGCCCG | TGTGGCCAGG | 720 |
| ACCATTCTGT | CAGCCTATGT | GGTCCTGAGG | CTGCCCTACC | AGCTGGCCCA | GCTGCTCTAC | 780 |
| CTGGCCTTCC | TGTGGGACGT | CTACTCTGGC | TACCTGCTCT | GGGAGGCCCT | GGTCTACTCC | 840 |
| GACTACCTGA | TCCTACTCAA | CAGCTGCCTC | AGCCCCTTCC | TCTGCCTCAT | GGCCAGTGCC | 900 |
| GACCTCCGGA | CCCTGCTGCG | CTCCGTGCTC | TCGTCCTTCG | CGGCAGCTCT | CTGCGAGGAG | 960 |
| CGGCCGGGCA | GCTTCACGCC | CACTGAGCCA | CAGACCCAGC | TAGATTCTGA | GGGTCCAACT | 1020 |
| CTGCCAGAGC | CGATGGCAGA | GGCCCAGTCA | CAGATGGATC | CTGTGGCCCA | GCCTCAGGTG | 1080 |
| AACCCCACAC | TCCAGCCACG | ATCGGATCCC | ACAGCTCAGC | CACAGCTGAA | CCCTACGGCC | 1140 |
| CAGCCACAGT | CGGATCCCAC | AGCCCAGCCA | CAGCTGAACC | TCATGGCCCA | GCCACAGTCA | 1200 |
| GATTCTGTGG | CCCAGCCACA | GGCAGACACT | AACGTCCAGA | CCCCTGCACC | TGCTGCCAGT | 1260 |
| TCTGTGCCCA | GTCCCTGTGA | TGAAGCTTCC | CCAACCCCAT | CCTCGCATCC | TACCCCAGGG | 1320 |
| GCCCTTGAGG | ACCCAGCCAC | ACCTCCTGCC | TCTGAAGGAG | AAAGCCCCAG | CAGCACCCCG | 1380 |
| CCAGAGGCGG | | AGGCCCCACG | TGA | | | 1413 |

(11) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 468 amino acids
 - (B) TYPE: amino acid
- (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Asp Thr Thr Met Glu Ala Asp Leu Gly Ala Thr Gly His Arg Pro 30 1 5 10 15

| | Arg | Thr | Glu | Leu 20 | Asp | Asp | Glu | Asp | Ser 25 | Tyr | Pro | Gln | Gly | Gly 30 | Trp | Asp |
|----|------------|------------|------------|------------|------------|------------|--------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | Thr | Val | Phe 35 | Leu | Val | Ala | Leu | Leu 40 | Leu | Leu | Gly | Leu | Pro 45 | Ala | Asn | Gly |
| 5 | Leu | Met 50 | Ala | Trp | Leu | Ala | Gly 55 | Ser | Gln | Ala | Arg | His 60 | Gly | Ala | Gly | Thr |
| | Arg 65 | Leu | Ala | Leu | Leu | Leu 70 | Leu | Ser | Leu | Ala | Leu 75 | Ser | Asp | Phe | Leu | Phe 80 |
| 10 | Leu | Ala | Ala | Ala | Ala 85 | Phe | Gln | Ile | Leu | Glu 90 | Ile | Arg | His | Gly | Gly 95 | His |
| | Trp | Pro | Leu | Gly 100 | Thr | Ala | Ala | Cys | Arg 105 | Phe | Tyr | Tyr | Phe | Leu 110 | Trp | Gly |
| | Val | Ser | Tyr 115 | Ser | Ser | Gly | Leu | Phe 120 | Leu | Leu | Ala | Ala | Leu 125 | Ser | Leu | Asp |
| 15 | Arg | Cys 130 | Leu | Leu | Ala | Leu | Cys 135 | Pro | His | Trp | Tyr | Pro 140 | Gly | His | Arg | Pro |
| | Val 145 | Arg | Leu | Pro | Leu | Trp 150 | Val | Cys | Ala | Gly | Val 155 | Trp | Val | Leu | Ala | Thr 160 |
| 20 | Leu | Phe | Ser | Val | Pro 165 | Trp | Leu | Val | Phe | Pro 170 | Glu | Ala | Ala | Val | Trp 175 | Trp |
| | Tyr | Asp | Leu | Val 180 | Ile | Cys | Leu | Asp | Phe 185 | | Asp | Ser | Glu | Glu 190 | Leu | Ser |
| | Leu | Arg | Met 195 | | Glu | Val | Leu | Gly 200 | | Phe | Leu | Pro | Phe 205 | | Leu | Leu |
| 25 | Leu | Val 210 | _ | His | Val | Leu | Thr 215 | | Ala | Thr | Arg | Thr 220 | | His | Arg | Gln |
| | Gln 225 | | Pro | Ala | Ala | Cys 230 | | Gly | Phe | : Ala | Arg 235 | | Ala | Arg | Thr | Ile 240 |
| 30 | Leu | Ser | Ala | Tyr | Val 245 | | Leu | Arg | Leu | 250 | | Gln | Leu | . Ala | Gln 255 | Leu |
| | Leu | Tyr | Leu | Ala 260 | | . Leu | Trp | Asp | Val 265 | | Ser | Gly | Tyr | 270 | | Trp |
| | Glu | Ala | Leu 275 | | Туг | Ser | Asp | Tyr 280 | | ı Ile | e Lev | Leu | Asr 285 | | . Càa | Leu |
| 35 | Ser | 290 | | e Leu | . Cys | Leu | 1 Met 295 | | ser | c Ala | a Asp | 300 | | g Thi | Leu | ı Leu |
| | Arg | g Ser | c Val | Let | Ser | : Sei | . Phe | Ala | a Ala | a Ala | a Lei | ı Cys | s Gli | ı Glı | ı Arg | g Pro |

| ARE | N-0054 | | | | | | - 7 | ' 5 - | PA | | | | TEN' | Т | | | |
|--|------------|------------|------------|------------|------------|------------|------------|--------------|------------|------------|------------|------------|------------|------------|------------|------------|-----|
| | 305 | | | | | 310 | | | | | 315 | | | | | 320 | |
| | Gly | Ser | Phe | Thr | Pro 325 | Thr | Glu | Pro | Gln | Thr 330 | Gln | Leu | Asp | Ser | Glu 335 | Gly | |
| 5 | Pro | Thr | Leu | Pro 340 | Glu | Pro | Met | Ala | Glu 345 | Ala | Gln | Ser | Gln | Met 350 | Asp | Pro | |
| | Val | Ala | Gln 355 | Pro | Gln | Val | Asn | Pro 360 | Thr | Leu | Gln | Pro | Arg 365 | Ser | Asp | Pro | |
| | Thr | Ala 370 | Gln | Pro | Gln | Leu | Asn 375 | Pro | Thr | Ala | Gln | Pro 380 | Gln | Ser | Asp | Pro | |
| 10 | Thr 385 | Ala | Gln | Pro | Gln | Leu 390 | Asn | Leu | Met | Ala | Gln 395 | Pro | Gln | Ser | Asp | Ser 400 | |
| | Val | Ala | Gln | Pro | Gln 405 | Ala | Asp | Thr | Asn | Val 410 | Gln | Thr | Pro | Ala | Pro 415 | Ala | |
| 15 | Ala | Ser | Ser | Val 420 | Pro | Ser | Pro | Cys | Asp 425 | Glu | Ala | Ser | Pro | Thr 430 | Pro | Ser | |
| | Ser | His | Pro 435 | Thr | Pro | Gly | Ala | Leu 440 | Glu | Asp | Pro | Ala | Thr 445 | Pro | Pro | Ala | |
| | Ser | Glu 450 | Gly | Glu | Ser | Pro | Ser 455 | | Thr | Pro | Pro | Glu 460 | Ala | Ala | Pro | Gly | |
| 20 | Ala 465 | Gly | Pro | Thr | | | | | | | | | | | | | |
| | (12) INF | ORMA | TION | FOR | SEQ | ID | NO:1 | 1: | | | | | | | | | |
| (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1248 base pairs 25 (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | | | | | | | | | | | | | | | | | |
| | (ii) | MOL | ECUL | E TY | PE: | DNA | (gen | omic | :) | | | | | | | | |
| | (xi) | SEQ | UENC | E DE | SCRI | PTIO | N: S | EQ I | D NO | :11: | | | | | | | |
| 30 | ATGTCAGG | GA I | 'GGAA | AAAC | T TC | AGAA | TGCT | TCC | TGGA | TCT | ACCA | .GCAG | AA A | .CTAG | AAGA | T | 60 |
| | CCATTCC | GA A | ACAC | CTGA | A CA | GCAC | CGAG | GAG | TATO | TGG | CCTI | CCTC | TG C | GGAC | CTCG | G | 120 |
| | CGCAGCCA | CT I | CTTC | CTCC | c ca | TGTC | TGTG | GTG | TATG | TGC | CAAT | TTTT | GT G | GTGG | GGGI | C | 180 |
| | ATTGGCA | TG I | CCTC | GTGT | G CC | TGGT | GATT | CTC | CAGO | ACC | AGGC | TATO | AA C | ACGO | CCAC | :C | 240 |

AACTACTACC TCTTCAGCCT GGCGGTCTCT GACCTCCTGG TCCTGCTCCT TGGAATGCCC 300

| ARE | N-0054 | | - 7 | 76 - | | PA | TENT |
|-----|------------|--|--------------------------|------------------|-------------|------------------|------|
| | CTGGAGGTCT | ATGAGATGTG | GCGCAACTAC | CCTTTCTTGT | TCGGGCCCGT | GGGCTGCTAC | 360 |
| | TTCAAGACGG | CCCTCTTTGA | GACCGTGTGC | TTCGCCTCCA | TCCTCAGCAT | CACCACCGTC | 420 |
| | AGCGTGGAGC | GCTACGTGGC | CATCCTACAC | CCGTTCCGCG | CCAAACTGCA | GAGCACCCGG | 480 |
| | CGCCGGGCCC | TCAGGATCCT | CGGCATCGTC | TGGGGCTTCT | CCGTGCTCTT | CTCCCTGCCC | 540 |
| 5 | AACACCAGCA | TCCATGGCAT | CAAGTTCCAC | TACTTCCCCA | ATGGGTCCCT | GGTCCCAGGT | 600 |
| | TCGGCCACCT | GTACGGTCAT | CAAGCCCATG | TGGATCTACA | ATTTCATCAT | CCAGGTCACC | 660 |
| | TCCTTCCTAT | TCTACCTCCT | CCCCATGACT | GTCATCAGTG | TCCTCTACTA | CCTCATGGCA | 720 |
| | CTCAGACTAA | AGAAAGACAA | ATCTCTTGAG | GCAGATGAAG | GGAATGCAAA | TATTCAAAGA | 780 |
| | CCCTGCAGAA | AATCAGTCAA | CAAGATGCTG | TTTGTCTTGG | TCTTAGTGTT | TGCTATCTGT | 840 |
| 10 | TGGGCCCCGT | TCCACATTGA | CCGACTCTTC | TTCAGCTTTG | TGGAGGAGTG | GAGTGAATCC | 900 |
| | CTGGCTGCTG | TGTTCAACCT | CGTCCATGTG | GTGTCAGGTG | TCTTCTTCTA | CCTGAGCTCA | 960 |
| | GCTGTCAACC | CCATTATCTA | TAACCTACTG | TCTCGCCGCT | TCCAGGCAGC | ATTCCAGAAT | 1020 |
| | GTGATCTCTT | CTTTCCACAA | ACAGTGGCAC | TCCCAGCATG | ACCCACAGTT | GCCACCTGCC | 1080 |
| | CAGCGGAACA | TCTTCCTGAC | AGAATGCCAC | TTTGTGGAGC | TGACCGAAGA | TATAGGTCCC | 1140 |
| 15 | CAATTCCCAT | GTCAGTCATC | CATGCACAAC | TCTCACCTCC | CAACAGCCCT | CTCTAGTGAA | 1200 |
| | CAGATGTCAA | GAACAAACTA | TCAAAGCTTC | CACTTTAACA | AAACCTGA | | 1248 |
| | (13) INFOR | MATION FOR | SEQ ID NO:1 | 2: | | | |
| 20 | | EQUENCE CHAR (A) LENGTH: (B) TYPE: au (C) STRANDEI (D) TOPOLOG | 415 amino a mino acid | acids | | | |
| | (ii) M | OLECULE TYP | E: protein | | | | |
| | (xi) S | EQUENCE DES | CRIPTION: S | EQ ID NO:12 | : | | |
| 25 | Met S 1 | Ser Gly Met | Glu Lys Leu 5 | Gln Asn Al 10 | | le Tyr Gln 15 | Gln |
| | Lys I | eu Glu Asp | Pro Phe Gln | Lys His Le | u Asn Ser T | hr Glu Glu | Tyr |

Leu Ala Phe Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val

Ser Val Val Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val

W. W.

IFI

- 78 -

360

Cys His Phe Val Glu Leu Thr Glu Asp Ile Gly Pro Gln Phe Pro Cys 375

5 Gln Ser Ser Met His Asn Ser His Leu Pro Thr Ala Leu Ser Ser Glu 390 395

> Gln Met Ser Arg Thr Asn Tyr Gln Ser Phe His Phe Asn Lys Thr 405 410

(14) INFORMATION FOR SEQ ID NO:13:

10 (i) SEQUENCE CHARACTERISTICS:

AREN-0054

ı 12

٠... m

II. II.

- (A) LENGTH: 1173 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ATGCCAGATA CTAATAGCAC AATCAATTTA TCACTAAGCA CTCGTGTTAC TTTAGCATTT 60 TTTATGTCCT TAGTAGCTTT TGCTATAATG CTAGGAAATG CTTTGGTCAT TTTAGCTTTT 120 GTGGTGGACA AAAACCTTAG ACATCGAAGT AGTTATTTTT TTCTTAACTT GGCCATCTCT 180 20 GACTTCTTG TGGGTGTGAT CTCCATTCCT TTGTACATCC CTCACACGCT GTTCGAATGG 240 GATTTTGGAA AGGAAATCTG TGTATTTTGG CTCACTACTG ACTATCTGTT ATGTACAGCA 300 TCTGTATATA ACATTGTCCT CATCAGCTAT GATCGATACC TGTCAGTCTC AAATGCTGTG 360 TCTTATAGAA CTCAACATAC TGGGGTCTTG AAGATTGTTA CTCTGATGGT GGCCGTTTGG 420 GTGCTGGCCT TCTTAGTGAA TGGGCCAATG ATTCTAGTTT CAGAGTCTTG GAAGGATGAA 480 25 GGTAGTGAAT GTGAACCTGG ATTTTTTCG GAATGGTACA TCCTTGCCAT CACATCATTC 540 TTGGAATTCG TGATCCCAGT CATCTTAGTC GCTTATTTCA ACATGAATAT TTATTGGAGC 600 CTGTGGAAGC GTGATCATCT CAGTAGGTGC CAAAGCCATC CTGGACTGAC TGCTGTCTCT 660 TCCAACATCT GTGGACACTC ATTCAGAGGT AGACTATCTT CAAGGAGATC TCTTTCTGCA 720 TCGACAGAAG TTCCTGCATC CTTTCATTCA GAGAGACAGA GGAGAAAGAG TAGTCTCATG 780 30 TTTTCCTCAA GAACCAAGAT GAATAGCAAT ACAATTGCTT CCAAAATGGG TTCCTTCTCC 840 CAATCAGATT CTGTAGCTCT TCACCAAAGG GAACATGTTG AACTGCTTAG AGCCAGGAGA 900

Leu Val Asn Gly Pro Met Ile Leu Val Ser Glu Ser Trp Lys Asp Glu

Gly Ser Glu Cys Glu Pro Gly Phe Phe Ser Glu Trp Tyr Ile Leu Ala

155

170

150

165

ij

IU

En will

The state of the s

| | | Ile | Thr | Ser | Phe 180 | Leu | Glu | Phe | Val | Ile 185 | Pro | Val | Ile | Leu | Val 190 | Ala | Tyr |
|----|------|------------|------------------------|----------------------------------|---------------------|---------------|--------------------|-----------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | | Phe | Asn | Met 195 | Asn | Ile | Tyr | Trp | Ser 200 | Leu | Trp | Lys | Arg | Asp 205 | His | Leu | Ser |
| 5 | | Arg | Cys 210 | Gln | Ser | His | Pro | Gly 215 | Leu | Thr | Ala | Val | Ser 220 | Ser | Asn | Ile | Cys |
| | | Gly 225 | His | Ser | Phe | Arg | Gly 230 | Arg | Leu | Ser | Ser | Arg 235 | Arg | Ser | Leu | Ser | Ala 240 |
| 10 | | Ser | Thr | Glu | Val | Pro 245 | Ala | Ser | Phe | His | Ser 250 | Glu | Arg | Gln | Arg | Arg 255 | Lys |
| | | Ser | Ser | Leu | Met 260 | Phe | Ser | Ser | Arg | Thr 265 | Lys | Met | Asn | Ser | Asn 270 | Thr | Ile |
| | | Ala | Ser | Lys 275 | Met | Gly | Ser | Phe | Ser 280 | Gln | Ser | Asp | Ser | Val 285 | Ala | Leu | His |
| 15 | | Gln | Arg 290 | Glu | His | Val | Glu | Leu 295 | Leu | Arg | Ala | Arg | Arg 300 | Leu | Ala | Lys | Seŗ |
| | | Leu 305 | Ala | Ile | Leu | Leu | Gly 310 | Val | Phe | Ala | Val | Cys 315 | Trp | Ala | Pro | Tyr | Ser 320 |
| 20 | | Leu | Phe | Thr | Ile | Val 325 | Leu | Ser | Phe | Tyr | Ser 330 | Ser | Ala | Thr | Gly | Pro 335 | Lys |
| | | Ser | Val | Trp | Tyr 340 | _ | Ile | Ala | Phe | Trp | | Gln | Trp | Phe | Asn 350 | | Phe |
| | | Val | Asn | Pro 355 | | Leu | Tyr | Pro | Leu 360 | | His | Lys | Arg | Phe 365 | | Lys | Ala |
| 25 | | Phe | Leu 370 | | Ile | Phe | Cys | Ile 375 | | Lys | Gln | Pro | Leu 380 | | Ser | Gln | His |
| | | Ser 385 | _ | Ser | · Val | . Ser | Ser 390 | | | | | | | | | | |
| | (16) | INF | 'ORMA | TION | FOR | SEÇ |) ID | NO:1 | .5 : | | | | | | | | |
| 30 | | (i) | (<i>P</i> (E (C | OUENC A) LE B) TY C) SI | NGTH PE: RANI | I: 30 nucl | bas eic ESS: | se pa acid sing | irs l | | | | | | | | |
| 35 | | (ii) | MOI | ECUI | E TY | PE: | DNA | (ger | omio | 2) | | | | | | | |

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

(iv) ANTI-SENSE: NO

15

25

30

How with

Ш

And the same of th

30

| (17) | INFORMATION | FOR | SEQ | ID | NO:16: |
|------|-------------|-----|-----|----|--------|

GGAAAGCTTA ACGATCCCCA GGAGCAACAT

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (iv) ANTI-SENSE: YES
- 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTGGGATCCT ACGAGAGCAT TTTTCACACA G

- (18) INFORMATION FOR SEQ ID NO:17:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1128 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

| ATGGCGAACG | CGAGCGAGCC | GGGTGGCAGC | GGCGGCGGCG | AGGCGGCCGC | CCTGGGCCTC | 60 |
|------------|------------|------------|------------|------------|------------|-----|
| AAGCTGGCCA | CGCTCAGCCT | GCTGCTGTGC | GTGAGCCTAG | CGGGCAACGT | GCTGTTCGCG | 120 |
| CTGCTGATCG | TGCGGGAGCG | CAGCCTGCAC | CGCGCCCCGT | ACTACCTGCT | GCTCGACCTG | 180 |
| TGCCTGGCCG | ACGGGCTGCG | CGCGCTCGCC | TGCCTCCCGG | CCGTCATGCT | GGCGGCGCGG | 240 |
| CGTGCGGCGG | ccgcggcgg | GGCGCCGCCG | GGCGCGCTGG | GCTGCAAGCT | GCTCGCCTTC | 300 |
| CTGGCCGCGC | TCTTCTGCTT | CCACGCCGCC | TTCCTGCTGC | TGGGCGTGGG | CGTCACCCGC | 360 |
| TACCTGGCCA | TCGCGCACCA | CCGCTTCTAT | GCAGAGCGCC | TGGCCGGCTG | GCCGTGCGCC | 420 |
| GCCATGCTGG | TGTGCGCCGC | CTGGGCGCTG | GCGCTGGCCG | CGGCCTTCCC | GCCAGTGCTG | 480 |
| GACGGCGGTG | GCGACGACGA | GGACGCGCCG | TGCGCCCTGG | AGCAGCGGCC | CGACGGCGCC | 540 |
| CCCGGCGCGC | TGGGCTTCCT | GCTGCTGCTG | GCCGTGGTGG | TGGGCGCCAC | GCACCTCGTC | 600 |
| TACCTCCGCC | TGCTCTTCTT | CATCCACGAC | CGCCGCAAGA | TGCGGCCCGC | GCGCCTGGTG | 660 |

| CCCGCCGTCA | GCCACGACTG | GACCTTCCAC | GGCCCGGGCG | CCACCGGCCA | GGCGGCCGCC | 720 |
|------------|------------|------------|------------|------------|------------|------|
| AACTGGACGG | CGGGCTTCGG | CCGCGGGCCC | ACGCCGCCCG | CGCTTGTGGG | CATCCGGCCC | 780 |
| GCAGGGCCGG | gccgcggcgc | GCGCCGCCTC | CTCGTGCTGG | AAGAATTCAA | GACGGAGAAG | 840 |
| AGGCTGTGCA | AGATGTTCTA | CGCCGTCACG | CTGCTCTTCC | TGCTCCTCTG | GGGGCCCTAC | 900 |
| GTCGTGGCCA | GCTACCTGCG | GGTCCTGGTG | CGGCCCGGCG | CCGTCCCCCA | GGCCTACCTG | 960 |
| ACGGCCTCCG | TGTGGCTGAC | CTTCGCGCAG | GCCGGCATCA | ACCCCGTCGT | GTGCTTCCTC | 1020 |
| TTCAACAGGG | AGCTGAGGGA | CTGCTTCAGG | GCCCAGTTCC | CCTGCTGCCA | GAGCCCCCGG | 1080 |
| ACCACCCAGG | CGACCCATCC | CTGCGACCTG | AAAGGCATTG | GTTTATGA | | 1128 |

- (19) INFORMATION FOR SEQ ID NO:18:
- 10 (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 375 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- 15 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Ala Asn Ala Ser Glu Pro Gly Gly Ser Gly Gly Gly Glu Ala Ala 1 5 10 15

Ala Leu Gly Leu Lys Leu Ala Thr Leu Ser Leu Leu Cys Val Ser 20 25 30

Leu Ala Gly Asn Val Leu Phe Ala Leu Leu Ile Val Arg Glu Arg Ser 35 40 45

Leu His Arg Ala Pro Tyr Tyr Leu Leu Leu Asp Leu Cys Leu Ala Asp 50 55 60

25 Gly Leu Arg Ala Leu Ala Cys Leu Pro Ala Val Met Leu Ala Ala Arg 65 70 75 80

> Arg Ala Ala Ala Ala Gly Ala Pro Pro Gly Ala Leu Gly Cys Lys 85 90 95

Leu Leu Ala Phe Leu Ala Ala Leu Phe Cys Phe His Ala Ala Phe Leu 30 100 105 110

Leu Leu Gly Val Gly Val Thr Arg Tyr Leu Ala Ile Ala His His Arg 115 120 125

Phe Tyr Ala Glu Arg Leu Ala Gly Trp Pro Cys Ala Ala Met Leu Val 130 135 140

| | | Cys 145 | Ala | Ala | Trp | Ala | Leu 150 | Ala | Leu | Ala | Ala | Ala 155 | Phe | Pro | Pro | Val | Leu 160 |
|----|------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | | Asp | Gly | Gly | Gly | Asp 165 | Asp | Glu | Asp | Ala | Pro 170 | Cys | Ala | Leu | Glu | Gln 175 | Arg |
| 5 | | Pro | Asp | Gly | Ala 180 | Pro | Gly | Ala | Leu | Gly 185 | Phe | Leu | Leu | Leu | Leu 190 | Ala | Val |
| | | Val | Val | Gly 195 | Ala | Thr | His | Leu | Val 200 | Tyr | Leu | Arg | Leu | Leu 205 | Phe | Phe | Ile |
| 10 | | His | Asp 210 | Arg | Arg | Lys | Met | Arg 215 | Pro | Ala | Arg | Leu | Val 220 | Pro | Ala | Val | Ser |
| | | His 225 | Asp | Trp | Thr | Phe | His 230 | Gly | Pro | Gly | Ala | Thr 235 | Gly | Gln | Ala | Ala | Ala 240 |
| | | Asn | Trp | Thr | Ala | Gly 245 | Phe | Gly | Arg | Gly | Pro 250 | Thr | Pro | Pro | Ala | Leu 255 | Val |
| 15 | | Gly | Ile | Arg | Pro 260 | Ala | Gly | Pro | Gly | Arg 265 | Gly | Ala | Arg | Arg | Leu 270 | Leu | Val |
| | | Leu | Glu | Glu 275 | Phe | Lys | Thr | Glu | Lys 280 | Arg | Leu | Cys | Lys | Met 285 | Phe | Tyr | Ala |
| 20 | | Val | Thr 290 | Leu | Leu | Phe | Leu | Leu 295 | Leu | Trp | Gly | Pro | Туг 300 | Val | Val | Ala | Ser |
| | | Tyr 305 | Leu | Arg | Val | Leu | Val 310 | Arg | Pro | Gly | Ala | Val 315 | | Gln | Ala | Tyr | Leu 320 |
| | | Thr | Ala | Ser | Val | Trp 325 | | Thr | Phe | Ala | Gln 330 | | Gly | Ile | Asn | Pro 335 | Val |
| 25 | | Val | Cys | Phe | Leu 340 | Phe | Asn | Arg | Glu | Leu 345 | | Asp | Cys | Phe | Arg 350 | | Gln |
| | | Phe | Pro | Cys 355 | | Gln | Ser | Pro | Arg 360 | | Thr | Gln | Ala | Thr 365 | | Pro | Cys |
| 30 | | Asp | Leu 370 | _ | Gly | Ile | Gly | Leu 375 | | | | | | | | | |
| | (20) | INF | ORMA | TION | FOR | SEC | ID | NO:1 | 9: | | | | | | | | |

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1002 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

| | ATGAACACCA | CAGTGATGCA | AGGCTTCAAC | AGATCTGAGC | GGTGCCCCAG | AGACACTCGG | 60 |
|----|---------------------------------|------------|--------------|------------|-------------|--------------|------|
| | ATAGTACAGC | TGGTATTCCC | AGCCCTCTAC | ACAGTGGTTT | TCTTGACCGG | CATCCTGCTG | 120 |
| | AATACTTTGG | CTCTGTGGGT | GTTTGTTCAC | ATCCCCAGCT | CCTCCACCTT | CATCATCTAC | 180 |
| 5 | CTCAAAAACA | CTTTGGTGGC | CGACTTGATA | ATGACACTCA | TGCTTCCTTT | CAAAATCCTC | 240 |
| | TCTGACTCAC | ACCTGGCACC | CTGGCAGCTC | AGAGCTTTTG | TGTGTCGTTT | TTCTTCGGTG | 300 |
| | ATATTTTATG | AGACCATGTA | TGTGGGCATC | GTGCTGTTAG | GGCTCATAGC | CTTTGACAGA | 360 |
| | TTCCTCAAGA | TCATCAGACC | TTTGAGAAAT | ATTTTTCTAA | AAAAACCTGT | TTTTGCAAAA | 420 |
| | ACGGTCTCAA | TCTTCATCTG | GTTCTTTTTG | TTCTTCATCT | CCCTGCCAAA | TACGATCTTG | 480 |
| 10 | AGCAACAAGG | AAGCAACACC | ATCGTCTGTG | AAAAAGTGTG | CTTCCTTAAA | GGGGCCTCTG | 540 |
| | GGGCTGAAAT | GGCATCAAAT | GGTAAATAAC | ATATGCCAGT | TTATTTTCTG | GACTGTTTTT | 600 |
| | ATCCTAATGC | TTGTGTTTTA | TGTGGTTATT | GCAAAAAAAG | TATATGATTC | TTATAGAAAG | ·660 |
| | TCCAAAAGTA | AGGACAGAAA | AAACAACAAA | AAGCTGGAAG | GCAAAGTATT | TGTTGTCGTG | 720 |
| | GCTGTCTTCT | TTGTGTGTTT | TGCTCCATTT | CATTTTGCCA | GAGTTCCATA | TACTCACAGT | 780 |
| 15 | CAAACCAACA | ATAAGACTGA | CTGTAGACTG | CAAAATCAAC | TGTTTATTGO | TAAAGAAACA | 840 |
| | ACTCTCTTT | TGGCAGCAAC | TAACATTTGT | arggarccc1 | TAATATACA | r ATTCTTATGT | 900 |
| | አ አ አ አ አ አ አ ጥጥ ሮ እ | СВСВВВВССТ | ' ACCATGTATO | CAAGGGAGA | A AGACCACAG | C ATCAAGCCAA | 960 |

- (21) INFORMATION FOR SEQ ID NO:20:
- 20 (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 333 amino acids

GAAAATCATA GCAGTCAGAC AGACAACATA ACCTTAGGCT GA

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- 25 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Asn Thr Thr Val Met Gln Gly Phe Asn Arg Ser Glu Arg Cys Pro

1002

Arg Asp Thr Arg Ile Val Gln Leu Val Phe Pro Ala Leu Tyr Thr Val 20 25 30

| | Val | Phe | Leu 35 | Thr | Gly | Ile | | Leu 40 | Asn | Thr | Leu | Ala | Leu 45 | Trp | Val | Phe |
|----|------------|------------|--------------|--------------|------------|------------|--------------|-------------|--------------|------------|------------|------------|--------------|--------------|--------------|--------------|
| | Val | His 50 | Ile | Pro | Ser | Ser | Ser 55 | Thr | Phe | Ile | Ile | Tyr 60 | Leu | Lys | Asn | Thr |
| 5 | Leu 65 | Val | Ala | Asp | Leu | Ile 70 | Met | Thr | Leu | Met | Leu 75 | Pro | Phe | Lys | Ile | Leu 80 |
| | Ser | Asp | Ser | His | Leu 85 | Ala | Pro | Trp | Gln | Leu 90 | Arg | Ala | Phe | Val | Cys 95 | Arg |
| 10 | Phe | Ser | Ser | Val 100 | Ile | Phe | Tyr | Glu | Thr 105 | Met | Tyr | Val | Gly | Ile 110 | Val | Leu |
| | Leu | Gly | Leu 115 | Ile | Ala | Phe | Asp | Arg 120 | Phe | Leu | Lys | Ile | Ile 125 | Arg | Pro | Leu |
| | Arg | Asn 130 | | Phe | Leu | Lys | Lys 135 | Pro | Val | Phe | Ala | Lys 140 | Thr | Val | Ser | Ile |
| 15 | Phe 145 | | Trp | Phe | Phe | Leu 150 | Phe | Phe | Ile | Ser | Leu 155 | Pro | Asn | Thr | Ile | Leu 160 |
| | Ser | Asn | . Lys | Glu | Ala 165 | | Pro | Ser | Ser | Val 170 | | Lys | Cys | Ala | Ser 175 | Leu |
| 20 | Lys | Gly | Pro | Leu 180 | | Leu | Lys | Trp | His 185 | | Met | Val | Asn | Asn 190 | Ile | Cys |
| | Gln | Phe | : Ile 195 | | Trp | Thr | Val | Phe 200 | | . Leu | Met | Leu | Val 205 | | Tyr | Val |
| | Val | . Ile | | a Lys | Lys | . Val | Tyr 215 | | Ser | туг | Arg | Lys 220 | | Lys | Ser | Lys |
| 25 | Asp 225 | | g Lys | s Asn | Asr | Lys 230 | | Leu | ı Glu | ı Gly | 235 | | . Phe | · Val | . Val | Val 240 |
| | Ala | a Val | l Phe | e Phe | 245 | | Phe | e Ala | e Pro | 250 | e His | : Phe | e Ala | Arg | 7 Val 255 | Pro |
| 30 | Ту | r Th: | r Hi: | s Ser 260 | | n Thi | . Asr | n Ası | 1 Ly: 26! | | r Asp | Cys | s Arg | J Let 270 | | n Asn |
| | Gli | n Le | u Pho 27 | | e Ala | a Lys | s Glu | 1 Th: 28 | | r Lei | u Phe | e Lei | u Ala 289 | a Ala | a Thi | Asn |
| | Il | e Cy 29 | | t As] | p Pr | o Lei | u Ile 29! | | r Il | e Ph | e Lei | з Су 30 | s Ly: 0 | s Ly: | s Phe | e Thr |
| 35 | G1 30 | | s Le | u Pr | э Су | s Me | | n Gl | y Ar | g Ly | s Th: | r Th 5 | r Al | a Se | r Se | r Gln 320 |
| | Gl | u As | n Hi | s Se | r Se | r Gl | n Th | r As | p As | n Il | e Th | r Le | u Gl | У | | |

(22) INFORMATION FOR SEQ ID NO:21:

5

the state of the s

100 mm 10

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1122 base pairs
- (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

| 10 | ATGGCCAACA | CTACCGGAGA | GCCTGAGGAG | GTGAGCGGCG | CTCTGTCCCC | ACCGTCCGCA | 60 |
|----|------------|------------|------------|------------|------------|------------|------|
| | TCAGCTTATG | TGAAGCTGGT | ACTGCTGGGA | CTGATTATGT | GCGTGAGCCT | GGCGGGTAAC | 120 |
| | GCCATCTTGT | CCCTGCTGGT | GCTCAAGGAG | CGTGCCCTGC | ACAAGGCTCC | TTACTACTTC | 180 |
| | CTGCTGGACC | TGTGCCTGGC | CGATGGCATA | CGCTCTGCCG | TCTGCTTCCC | CTTTGTGCTG | .240 |
| | GCTTCTGTGC | GCCACGGCTC | TTCATGGACC | TTCAGTGCAC | TCAGCTGCAA | GATTGTGGCC | 300 |
| 15 | TTTATGGCCG | TGCTCTTTTG | CTTCCATGCG | GCCTTCATGC | TGTTCTGCAT | CAGCGTCACC | 360 |
| | CGCTACATGG | CCATCGCCCA | CCACCGCTTC | TACGCCAAGC | GCATGACACT | CTGGACATGC | 420 |
| | GCGGCTGTCA | TCTGCATGGC | CTGGACCCTG | TCTGTGGCCA | TGGCCTTCCC | ACCTGTCTTT | 480 |
| | GACGTGGGCA | CCTACAAGTT | TATTCGGGAG | GAGGACCAGT | GCATCTTTGA | GCATCGCTAC | 540 |
| | TTCAAGGCCA | ATGACACGCT | GGGCTTCATG | CTTATGTTGG | CTGTGCTCAT | GGCAGCTACC | 600 |
| 20 | CATGCTGTCT | ACGGCAAGCT | GCTCCTCTTC | GAGTATCGTC | ACCGCAAGAT | GAAGCCAGTG | 660 |
| | CAGATGGTGC | CAGCCATCAG | CCAGAACTGG | ACATTCCATG | GTCCCGGGGC | CACCGGCCAG | 720 |
| | GCTGCTGCCA | ACTGGATCGC | CGGCTTTGGC | CGTGGGCCCA | TGCCACCAAC | CCTGCTGGGT | 780 |
| | ATCCGGCAGA | ATGGGCATGC | AGCCAGCCGG | CGGCTACTGG | GCATGGACGA | GGTCAAGGGT | 840 |
| | GAAAAGCAGC | TGGGCCGCAT | GTTCTACGCG | ATCACACTGC | TCTTTCTGCT | CCTCTGGTCA | 900 |
| 25 | CCCTACATCG | TGGCCTGCTA | CTGGCGAGTG | TTTGTGAAAG | CCTGTGCTGT | GCCCCACCGC | 960 |
| | TACCTGGCCA | CTGCTGTTTG | GATGAGCTTC | GCCCAGGCTG | CCGTCAACCC | AATTGTCTGC | 1020 |
| | TTCCTGCTCA | ACAAGGACCT | CAAGAAGTGC | CTGACCACTC | ACGCCCCCTG | CTGGGGCACA | 1080 |
| | GGAGGTGCCC | CGGCTCCCAG | AGAACCCTAC | TGTGTCATGT | ' GA | | 1122 |

(23) INFORMATION FOR SEQ ID NO:22:

| # # # # # # # # # # # # # # # # # # # | 1 |
|--|--------|
| 4 1 | 7 |
| 14 | 1 |
| + + | in the |
| 17 | 6 |
| HILL HILL HILL HILL HILL HILL HILL HILL | 1 |
| IJ | 200 |
| 11, | 7 |
| ä | |
| | 1 |
| Story. | 1 |
| 12 | 1 |
| ÷ + | il in |
| 4 8 | S S |
| 1 | i. |
| | |

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 373 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: 5 (D) TOPOLOGY: not relevant (ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22: Met Ala Asn Thr Thr Gly Glu Pro Glu Glu Val Ser Gly Ala Leu Ser Pro Pro Ser Ala Ser Ala Tyr Val Lys Leu Val Leu Leu Gly Leu Ile 10 25 Met Cys Val Ser Leu Ala Gly Asn Ala Ile Leu Ser Leu Leu Val Leu Lys Glu Arg Ala Leu His Lys Ala Pro Tyr Tyr Phe Leu Leu Asp Leu 15 Cys Leu Ala Asp Gly Ile Arg Ser Ala Val Cys Phe Pro Phe Val Leu Ala Ser Val Arg His Gly Ser Ser Trp Thr Phe Ser Ala Leu Ser Cys Lys Ile Val Ala Phe Met Ala Val Leu Phe Cys Phe His Ala Ala Phe 20 105 Met Leu Phe Cys Ile Ser Val Thr Arg Tyr Met Ala Ile Ala His His 120 Arg Phe Tyr Ala Lys Arg Met Thr Leu Trp Thr Cys Ala Ala Val Ile 25 Cys Met Ala Trp Thr Leu Ser Val Ala Met Ala Phe Pro Pro Val Phe 155 150 Asp Val Gly Thr Tyr Lys Phe Ile Arg Glu Glu Asp Gln Cys Ile Phe Glu His Arg Tyr Phe Lys Ala Asn Asp Thr Leu Gly Phe Met Leu Met 30 185 Leu Ala Val Leu Met Ala Ala Thr His Ala Val Tyr Gly Lys Leu Leu 200 195 Leu Phe Glu Tyr Arg His Arg Lys Met Lys Pro Val Gln Met Val Pro 35 215 Ala Ile Ser Gln Asn Trp Thr Phe His Gly Pro Gly Ala Thr Gly Gln 240 230 225

| -11111 | 22 | |
|---------|------------|---------|
| ä | 1 | 1 |
| | Ĭ | |
| ÷ | i. | 1 |
| ·mii. | 1 | |
| mun | 1 | 1 |
| пене | Sire. | 22.00 |
| out | 1 | feftit. |
| 5 | | |
| 31111 | 111 | |
| .mm. | # | |
| Time. | | |
| ÷ | ì. | 27.00 |
| .11331. | 122 122 | |
| 1111111 | 442 | = |

| | | Ala | Ala | Ala | Asn | Trp 245 | Ile | Ala | Gly | Phe | Gly 250 | Arg | Gly | Pro | Met | Pro 255 | Pro |
|----|------|------------|----------------|-----------------------|-----------------------------|----------------------|--------------------|---------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | | Thr | Leu | Leu | Gly 260 | Ile | Arg | Gln | Asn | Gly 265 | His | Ala | Ala | Ser | Arg 270 | Arg | Leu |
| 5 | | Leu | Gly | Met 275 | Asp | Glu | Val | Lys | Gly 280 | Glu | Lys | Gln | Leu | Gly 285 | Arg | Met | Phe |
| | | Tyr | Ala 290 | Ile | Thr | Leu | Leu | Phe 295 | Leu | Leu | Leu | Trp | Ser 300 | Pro | Tyr | Ile | Val |
| 10 | | Ala 305 | Cys | Tyr | Trp | Arg | Val 310 | Phe | Val | Lys | Ala | Cys 315 | Ala | Val | Pro | His | Arg 320 |
| | | Tyr | Leu | Ala | Thr | Ala 325 | Val | Trp | Met | Ser | Phe 330 | Ala | Gln | Ala | Ala | Val 335 | Asn |
| | | Pro | Ile | Val | Cys 340 | Phe | Leu | Leu | Asn | Lys 345 | | Leu | Lys | Lys | Cys 350 | Leu | Thr |
| 15 | | Thr | His | Ala 355 | Pro | Cys | Trp | Gly | Thr 360 | Gly | Gly | Ala | Pro | Ala 365 | | Arg | Glų |
| | | Pro | Tyr 370 | - | Val | Met | | | | | | | | | | | |
| | (24) | INF | ORMA | TION | FOR | SEQ | ID | NO:2 | 3: | | | | | | | | |
| 20 | | (i) | (A (B (C |) LE) TY !) ST | E CH NGTH PE: RAND | : 10 nucl EDNE | 53 b eic SS: | ase acid sing | pair | s | | | | | | | |
| 25 | | (ii) | MOL | ECUL | E TY | PE: | DNA | (ger | omic | :) | | | | | | | |

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

ATGGCTTTGG AACAGAACCA GTCAACAGAT TATTATTATG AGGAAAATGA AATGAATGGC 60

ACTTATGACT ACAGTCAATA TGAATTGATC TGTATCAAAG AAGATGTCAG AGAATTTGCA 120

AAAGTTTTCC TCCCTGTATT CCTCACAATA GCTTTCGTCA TTGGACTTGC AGGCAATTCC 180

30 ATGGTAGTGG CAATTTATGC CTATTACAAG AAACAGAGAA CCAAAACAGA TGTGTACATC 240

CTGAATTTGG CTGTAGCAGA TTTACTCCTT CTATTCACTC TGCCTTTTTG GGCTGTTAAT 300

GCAGTTCATG GGTGGGTTTT AGGGAAAATA ATGTGCAAAA TAACTTCAGC CTTGTACACA 360

CTAAACTTTG TCTCTGGAAT GCAGTTTCTG GCTTGCATCA GCATAGACAG ATATGTGGCA 420

GTAACTAATG TCCCCAGCCA ATCAGGAGTG GGAAAACCAT GCTGGATCAT CTGTTTCTGT 480

| ARE | N-0054 | | | - 89 |) _ | | | | | PAT | ENT |
|-----|------------|--|---|---------------------------------|-------------|--------------|----------|-------------|-----------|-----------|-----------|
| | GTCTGGATGG | G CTGCCATC | TT GCTGA | GCATA (| CCCAG | CTGG T | TTTTTAT. | AC AG | TAAAT | GAC | 540 |
| | AATGCTAGGT | r GCATTCCC | TTTCC | cccgc : | raccta(| GGAA C | ATCAATG | AA AG | CATTO | TTA | 600 |
| | CAAATGCTAG | G AGATCTGC | AT TGGAT | TTGTA (| GTACCC' | TTTC T | TATTATG | GG GG | rgtgo | CTAC | 660 |
| | TTTATCACGO | G CAAGGACA | CT CATGA | AGATG (| CCAAAC | ATTA A | AATATCT | CG AC | CCCT | AAAA | 720 |
| 5 | GTTCTGCTC | A CAGTCGTT | AT AGTTT | TCATT | GTCACT | CAAC T | GCCTTAT | AA CA | TTGT | CAAG | 780 |
| | TTCTGCCGA | G CCATAGAC | AT CATCT | ACTCC | CTGATC | ACCA G | CTGCAAC | AT GA | GCAA | ACGC | 840 |
| | ATGGACATC | G CCATCCAA | GT CACAG | BAAAGC | ATTGCA | CTCT I | TCACAGO | TG CC | TCAA | CCCA | 900 |
| | ATCCTTTAT | G TTTTTATG | GG AGCAT | CTTTC | AAAAAC | TACG T | TATGAA | GT GG | CCAA | GAAA | 960 |
| | TATGGGTCC | T GGAGAAGA | CA GAGAC | CAAAGT | GTGGAG | GAGT 1 | TCCTTT | GA TI | 'CTGA | GGGT | 1020 |
| 10 | CCTACAGAG | C CAACCAGT | AC TTTT | AGCATT | TAA | | | | | | 1053 |
| | (25) INFO | RMATION FO | R SEQ II | NO:24 | :: | | | | | | |
| 15 | | SEQUENCE ((A) LENGT (B) TYPE (C) STRAI (D) TOPO | TH: 350 a amino a IDEDNESS LOGY: not | amino a acid : t relev | cids | | | | | | |
| | (11) | MOLECULE ' | .ipe: pr | Ocein | | | | | | | |
| | (xi) | SEQUENCE | ESCRIPT | ION: SI | EQ ID 1 | NO:24: | | | | | |
| 20 | Met 1 | Ala Leu G | lu Gln A 5 | sn Gln | Ser T | nr Asp 10 | Tyr Ty | r Tyr | Glu | Glu 15 | Asn |
| | Glu | Met Asn G | | yr Asp | Tyr S | | Tyr Gl | u Leu | Ile 30 | Cys | Ile |
| | Lys | Glu Asp V 35 | al Arg G | lu Phe | Ala L 40 | ys Val | . Phe Le | u Pro 45 | Val | Phe | Leu |
| 25 | Thr | Ile Ala P 50 | he Val I | le Gly 55 | Leu A | la Gly | Asn Se | | Val | Val | Ala |
| | Ile 65 | Tyr Ala T | | Lys Lys 70 | Gln A | rg Thi | r Lys Th | ır Asp | Val | Tyr | Ile 80 |

Leu Asn Leu Ala Val Ala Asp Leu Leu Leu Phe Thr Leu Pro Phe

Trp Ala Val Asn Ala Val His Gly Trp Val Leu Gly Lys Ile Met Cys 105

Lys Ile Thr Ser Ala Leu Tyr Thr Leu Asn Phe Val Ser Gly Met Gln

100

ij

H

1.1

F

H

HIS HILL HIS THE

120 125 115 Phe Leu Ala Cys Ile Ser Ile Asp Arg Tyr Val Ala Val Thr Asn Val 130 135 Pro Ser Gln Ser Gly Val Gly Lys Pro Cys Trp Ile Ile Cys Phe Cys 155 150 Val Trp Met Ala Ala Ile Leu Leu Ser Ile Pro Gln Leu Val Phe Tyr 170 165 Thr Val Asn Asp Asn Ala Arg Cys Ile Pro Ile Phe Pro Arg Tyr Leu 185 Gly Thr Ser Met Lys Ala Leu Ile Gln Met Leu Glu Ile Cys Ile Gly 10 200 Phe Val Val Pro Phe Leu Ile Met Gly Val Cys Tyr Phe Ile Thr Ala 215 Arg Thr Leu Met Lys Met Pro Asn Ile Lys Ile Ser Arg Pro Leu Lys 230 15 Val Leu Leu Thr Val Val Ile Val Phe Ile Val Thr Gln Leu Pro Tyr 245 Asn Ile Val Lys Phe Cys Arg Ala Ile Asp Ile Ile Tyr Ser Leu Ile 265 Thr Ser Cys Asn Met Ser Lys Arg Met Asp Ile Ala Ile Gln Val Thr 20 280 Glu Ser Ile Ala Leu Phe His Ser Cys Leu Asn Pro Ile Leu Tyr Val 295 290 Phe Met Gly Ala Ser Phe Lys Asn Tyr Val Met Lys Val Ala Lys Lys 310 25 Tyr Gly Ser Trp Arg Arg Gln Arg Gln Ser Val Glu Glu Phe Pro Phe 325 Asp Ser Glu Gly Pro Thr Glu Pro Thr Ser Thr Phe Ser Ile 345 (26) INFORMATION FOR SEQ ID NO:25: 30 (i) SEQUENCE CHARACTERISTICS:

- - (A) LENGTH: 1116 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear 35
 - (ii) MOLECULE TYPE: DNA (genomic)

AREN-0054

| (vi) | CECTENCE | DESCRIPTION: | SEO | TD | NO - 25 - |
|------|-----------|--------------|-----|-------|-----------|
| LX1 | SECULENCE | DESCRIPTIONS | 200 | 1.1.1 | NU:ZD: |

| | ATGCCAGGAA | ACGCCACCCC | AGTGACCACC | ACTGCCCCGT | GGGCCTCCCT | GGGCCTCTCC | 60 |
|----|------------|------------|------------|------------|------------|------------|------|
| | GCCAAGACCT | GCAACAACGT | GTCCTTCGAA | GAGAGCAGGA | TAGTCCTGGT | CGTGGTGTAC | 120 |
| | AGCGCGGTGT | GCACGCTGGG | GGTGCCGGCC | AACTGCCTGA | CTGCGTGGCT | GGCGCTGCTG | 180 |
| 5 | CAGGTACTGC | AGGGCAACGT | GCTGGCCGTC | TACCTGCTCT | GCCTGGCACT | CTGCGAACTG | 240 |
| | CTGTACACAG | GCACGCTGCC | ACTCTGGGTC | ATCTATATCC | GCAACCAGCA | CCGCTGGACC | 300 |
| | CTAGGCCTGC | TGGCCTCGAA | GGTGACCGCC | TACATCTTCT | TCTGCAACAT | CTACGTCAGC | 360 |
| | ATCCTCTTCC | TGTGCTGCAT | CTCCTGCGAC | CGCTTCGTGG | CCGTGGTGTA | CGCGCTGGAG | 420 |
| | AGTCGGGGCC | GCCGCCGCCG | GAGGACCGCC | ATCCTCATCT | CCGCCTGCAT | CTTCATCCTC | 480 |
| 10 | GTCGGGATCG | TTCACTACCC | GGTGTTCCAG | ACGGAAGACA | AGGAGACCTG | CTTTGACATG | 540 |
| | CTGCAGATGG | ACAGCAGGAT | TGCCGGGTAC | TACTACGCCA | GGTTCACCGT | TGGCTTTGCC | 600 |
| | ATCCCTCTCT | CCATCATCGC | CTTCACCAAC | CACCGGATTT | TCAGGAGCAT | CAAGCAGAGC | 660 |
| | ATGGGCTTAA | GCGCTGCCCA | GAAGGCCAAG | GTGAAGCACT | CGGCCATCGC | GGTGGTTGTC | 720 |
| | ATCTTCCTAG | TCTGCTTCGC | CCCGTACCAC | CTGGTTCTCC | TCGTCAAAGC | CGCTGCCTTT | 780 |
| 15 | TCCTACTACA | GAGGAGACAG | GAACGCCATG | TGCGGCTTGG | AGGAAAGGCT | GTACACAGCC | 840 |
| | TCTGTGGTGT | TTCTGTGCCT | GTCCACGGTG | AACGGCGTGG | CTGACCCCAT | TATCTACGTG | 900 |
| | CTGGCCACGG | ACCATTCCCG | CCAAGAAGTG | TCCAGAATCC | ATAAGGGGTG | GAAAGAGTGG | 960 |
| | TCCATGAAGA | CAGACGTCAC | CAGGCTCACC | CACAGCAGGG | ACACCGAGGA | GCTGCAGTCG | 1020 |
| | CCCGTGGCCC | TTGCAGACCA | CTACACCTTC | TCCAGGCCCG | TGCACCCACC | AGGGTCACCA | 1080 |
| 20 | TGCCCTGCAA | AGAGGCTGAT | TGAGGAGTCC | TGCTGA | | | 1116 |

(28) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 371 amino acids
 - (B) TYPE: amino acid
- (C) STRANDEDNESS:

25

- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Met Pro Gly Asn Ala Thr Pro Val Thr Thr Thr Ala Pro Trp Ala Ser 30 1 5 10 15

| | Leu | Gly | Leu | Ser 20 | Ala | Lys | Thr | Cys | Asn 25 | Asn | Val | Ser | Phe | Glu 30 | Glu | Ser |
|----|------------|------------|--------------|------------|------------|------------|--------------|------------|--------------|------------|------------|-------------|------------|------------|------------|------------------|
| | Arg | Ile | Val 35 | Leu | Val | Val | Val | Tyr 40 | Ser | Ala | Val | Cys | Thr 45 | Leu | Gly | Val |
| 5 | Pro | Ala 50 | Asn | Cys | Leu | Thr | Ala 55 | Trp | Leu | Ala | Leu | Leu 60 | Gln | Val | Leu | Gln |
| | Gly 65 | Asn | Val | Leu | Ala | Val 70 | Tyr | Leu | Leu | Cys | Leu 75 | Ala | Leu | Cys | Glu | Leu 80 |
| 10 | Leu | Tyr | Thr | Gly | Thr 85 | Leu | Pro | Leu | Trp | Val 90 | Ile | Tyr | Ile | Arg | Asn 95 | Gln |
| | His | Arg | Trp | Thr 100 | Leu | Gly | Leu | Leu | Ala 105 | Ser | Lys | Val | Thr | Ala 110 | Tyr | Ile |
| | Phe | Phe | Cys 115 | Asn | Ile | Tyr | Val | Ser 120 | Ile | Leu | Phe | Leu | Cys 125 | Cys | Ile | Ser |
| 15 | Cys | Asp 130 | Arg | Phe | Val | Ala | Val 135 | Val | Tyr | Ala | Leu | Glu 140 | Ser | Arg | Gly | Arg [.] |
| | Arg 145 | Arg | Arg | Arg | Thr | Ala 150 | Ile | Leu | Ile | Ser | Ala 155 | Cys | Ile | Phe | Ile | Leu 160 |
| 20 | Val | Gly | Ile | Val | His 165 | | Pro | Val | Phe | Gln 170 | | Glu | Asp | Lys | Glu 175 | Thr |
| | Cys | Phe | Asp | Met 180 | Leu | Gln | Met | Asp | Ser 185 | | Ile | Ala | Gly | Tyr 190 | Tyr | Tyr |
| | Ala | Arg | Phe 195 | | Val | Gly | Phe | Ala 200 | | Pro | Leu | Ser | 1le 205 | | Ala | Phe |
| 25 | Thr | Asn 210 | | Arg | Ile | Phe | Arg 215 | | Ile | Lys | Gln | Ser 220 | | Gly | Leu | Ser |
| | Ala 225 | | Gln | Lys | Ala | Lys 230 | | Lys | His | Ser | 235 | | e Ala | Val | . Val | Val 240 |
| 30 | Ile | Phe | . Lev | ı Val | Cys 245 | | : Ala | Pro | туг | His 250 | | . Val | L Leu | ı Lev | Val 255 | Lys |
| | Ala | a Ala | a Ala | Phe 260 | | Tyr | туг | Arg | g Gly 265 | | Arg | , Ası | n Ala | 270 | | Gly |
| | Lev | ı Glı | 1 Glu 275 | | , Le | і Туг | Thr | Ala 280 | | r Val | l Val | . Phe | 285 | | Leu | . Ser |
| 35 | Thi | 290 | | ı Gly | v Vai | l Ala | a Asp 295 | | o Ile | e Ile | ∋ Туг | 7 Va 300 | | ı Ala | a Thi | Asp |

His Ser Arg Gln Glu Val Ser Arg Ile His Lys Gly Trp Lys Glu Trp 320

Ser Met Lys Thr Asp Val Thr Arg Leu Thr His Ser Arg Asp Thr Glu 335

Glu Leu Gln Ser Pro Val Ala Leu Ala Asp His Tyr Thr Phe Ser Arg Arg Asp Arg Ser Arg 350

Pro Val His 355

Pro Pro Gly Ser Pro Cys Pro Ala Lys Arg Leu Ile Glu 360

Glu Ser Cys 10 370

5

15

(28) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1113 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

60 ATGGCGAACT ATAGCCATGC AGCTGACAAC ATTTTGCAAA ATCTCTCGCC TCTAACAGCC 20 TTTCTGAAAC TGACTTCCTT GGGTTTCATA ATAGGAGTCA GCGTGGTGGG CAACCTCCTG 120 ATCTCCATTT TGCTAGTGAA AGATAAGACC TTGCATAGAG CACCTTACTA CTTCCTGTTG 180 GATCTTTGCT GTTCAGATAT CCTCAGATCT GCAATTTGTT TCCCATTTGT GTTCAACTCT 240 GTCAAAAATG GCTCTACCTG GACTTATGGG ACTCTGACTT GCAAAGTGAT TGCCTTTCTG 300 GGGGTTTTGT CCTGTTTCCA CACTGCTTTC ATGCTCTTCT GCATCAGTGT CACCAGATAC 360 25 TTAGCTATCG CCCATCACCG CTTCTATACA AAGAGGCTGA CCTTTTGGAC GTGTCTGGCT 420 GTGATCTGTA TGGTGTGGAC TCTGTCTGTG GCCATGGCAT TTCCCCCGGT TTTAGACGTG 480 GGCACTTACT CATTCATTAG GGAGGAAGAT CAATGCACCT TCCAACACCG CTCCTTCAGG 540 GCTAATGATT CCTTAGGATT TATGCTGCTT CTTGCTCTCA TCCTCCTAGC CACACAGCTT 600 GTCTACCTCA AGCTGATATT TTTCGTCCAC GATCGAAGAA AAATGAAGCC AGTCCAGTTT 660 30 GTAGCAGCAG TCAGCCAGAA CTGGACTTTT CATGGTCCTG GAGCCAGTGG CCAGGCAGCT 720 GCCAATTGGC TAGCAGGATT TGGAAGGGGT CCCACACCAC CCACCTTGCT GGGCATCAGG 780 CAAAATGCAA ACACCACAGG CAGAAGAAGG CTATTGGTCT TAGACGAGTT CAAAATGGAG 840

| .mm. | |
|--------------------|------------------|
| á | C |
| ·uur | H |
| ÷ | t _e i |
| | Tri. |
| THE REAL PROPERTY. | |
| | F |
| | Ļ |
| 3 | |
| .:: | mij mij |
| 4 | Į. |
| 4000 | 777 <u>1</u> |
| ÷ | الم الم |
| 4000 | 1247 |
| 1144411 | |
| | |

| AAAAGAATCA | GCAGAATGTT | CTATATAATG | ACTTTTCTGT | TTCTAACCTT | GTGGGGCCCC | 900 |
|------------|------------|------------|------------|------------|------------|------|
| TACCTGGTGG | CCTGTTATTG | GAGAGTTTTT | GCAAGAGGGC | CTGTAGTACC | AGGGGGATTT | 960 |
| CTAACAGCTG | CTGTCTGGAT | GAGTTTTGCC | CAAGCAGGAA | TCAATCCTTT | TGTCTGCATT | 1020 |
| TTCTCAAACA | GGGAGCTGAG | GCGCTGTTTC | AGCACAACCC | TTCTTTACTG | CAGAAAATCC | 1080 |
| AGGTTACCAA | GGGAACCTTA | CTGTGTTATA | TGA | | | 1113 |

- (29) INFORMATION FOR SEQ ID NO:28:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 370 amino acids
 - (B) TYPE: amino acid
- 10

- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:
- Met Ala Asn Tyr Ser His Ala Ala Asp Asn Ile Leu Gln Asn Leu Ser
 15 1 5 10 15
 - Pro Leu Thr Ala Phe Leu Lys Leu Thr Ser Leu Gly Phe Ile Ile Gly 20 25 30
 - Val Ser Val Val Gly Asn Leu Leu Ile Ser Ile Leu Leu Val Lys Asp 35 40 45
- 20 Lys Thr Leu His Arg Ala Pro Tyr Tyr Phe Leu Leu Asp Leu Cys Cys 50 55 60
 - Ser Asp Ile Leu Arg Ser Ala Ile Cys Phe Pro Phe Val Phe Asn Ser 65 70 75 80
- Val Lys Asn Gly Ser Thr Trp Thr Tyr Gly Thr Leu Thr Cys Lys Val
 25 85 90 95
 - Ile Ala Phe Leu Gly Val Leu Ser Cys Phe His Thr Ala Phe Met Leu 100 105 110
 - Phe Cys Ile Ser Val Thr Arg Tyr Leu Ala Ile Ala His His Arg Phe 115 120 125
- Tyr Thr Lys Arg Leu Thr Phe Trp Thr Cys Leu Ala Val Ile Cys Met 130 135 140
 - Val Trp Thr Leu Ser Val Ala Met Ala Phe Pro Pro Val Leu Asp Val 145 150 150 160
 - Gly Thr Tyr Ser Phe Ile Arg Glu Glu Asp Gln Cys Thr Phe Gln His

ıO m

1.

T

171

The Tark Man and the same and t

- Arg Ser Phe Arg Ala Asn Asp Ser Leu Gly Phe Met Leu Leu Leu Ala 180 185
- Leu Ile Leu Leu Ala Thr Gln Leu Val Tyr Leu Lys Leu Ile Phe Phe 200
 - Val His Asp Arg Arg Lys Met Lys Pro Val Gln Phe Val Ala Ala Val 215 210
 - Ser Gln Asn Trp Thr Phe His Gly Pro Gly Ala Ser Gly Gln Ala Ala 230
- Ala Asn Trp Leu Ala Gly Phe Gly Arg Gly Pro Thr Pro Pro Thr Leu 10
 - Leu Gly Ile Arg Gln Asn Ala Asn Thr Thr Gly Arg Arg Leu Leu 265
- Val Leu Asp Glu Phe Lys Met Glu Lys Arg Ile Ser Arg Met Phe Tyr 15 280
 - Ile Met Thr Phe Leu Phe Leu Thr Leu Trp Gly Pro Tyr Leu Val Ala 295 290
 - Cys Tyr Trp Arg Val Phe Ala Arg Gly Pro Val Val Pro Gly Gly Phe
- Leu Thr Ala Ala Val Trp Met Ser Phe Ala Gln Ala Gly Ile Asn Pro 20 330
 - Phe Val Cys Ile Phe Ser Asn Arg Glu Leu Arg Arg Cys Phe Ser Thr 340 345
- Thr Leu Leu Tyr Cys Arg Lys Ser Arg Leu Pro Arg Glu Pro Tyr Cys 360 25

Val Ile 370

- (30) INFORMATION FOR SEQ ID NO:29:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1080 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29: 35

| | GCGATCGCGG | TGGCCCTGCC | CGTGGTGTAC | TCGCTGGTGG | CGGCGGTCAG | CATCCCGGGC | 120 |
|----|-------------|------------|------------|------------|------------|------------|------|
| | AACCTCTTCT | CTCTGTGGGT | GCTGTGCCGG | CGCATGGGGC | CCAGATCCCC | GTCGGTCATC | 180 |
| | TTCATGATCA | ACCTGAGCGT | CACGGACCTG | ATGCTGGCCA | GCGTGTTGCC | TTTCCAAATC | 240 |
| | TACTACCATT | GCAACCGCCA | CCACTGGGTA | TTCGGGGTGC | TGCTTTGCAA | CGTGGTGACC | 300 |
| 5 | GTGGCCTTTT | ACGCAAACAT | GTATTCCAGC | ATCCTCACCA | TGACCTGTAT | CAGCGTGGAG | 360 |
| | CGCTTCCTGG | GGGTCCTGTA | CCCGCTCAGC | TCCAAGCGCT | GGCGCCGCCG | TCGTTACGCG | 420 |
| | GTGGCCGCGT | GTGCAGGGAC | CTGGCTGCTG | CTCCTGACCG | CCCTGTGCCC | GCTGGCGCGC | 480 |
| | ACCGATCTCA | CCTACCCGGT | GCACGCCCTG | GGCATCATCA | CCTGCTTCGA | CGTCCTCAAG | 540 |
| | TGGACGATGC | TCCCCAGCGT | GGCCATGTGG | GCCGTGTTCC | TCTTCACCAT | CTTCATCCTG | 600 |
| 10 | CTGTTCCTCA | TCCCGTTCGT | GATCACCGTG | GCTTGTTACA | CGGCCACCAT | CCTCAAGCTG | 660 |
| | TTGCGCACGG | AGGAGGCGCA | CGGCCGGGAG | CAGCGGAGGC | GCGCGGTGGG | CCTGGCCGCG | 720 |
| | GTGGTCTTGC | TGGCCTTTGT | CACCTGCTTC | GCCCCAACA | ACTTCGTGCT | CCTGGCGCAC | 780 |
| | ATCGTGAGCC | GCCTGTTCTA | CGGCAAGAGC | TACTACCACG | TGTACAAGCT | CACGCTGTGT | 840 |
| | CTCAGCTGCC | TCAACAACTG | TCTGGACCCG | TTTGTTTATT | ACTTTGCGTC | CCGGGAATTC | 900 |
| 15 | CAGCTGCGCC | TGCGGGAATA | TTTGGGCTGC | CGCCGGGTGC | CCAGAGACAC | CCTGGACACG | 960 |
| | CGCCGCGAGA | GCCTCTTCTC | CGCCAGGACC | ACGTCCGTGC | GCTCCGAGGC | CGGTGCGCAC | 1020 |
| | ССТСА АСССА | TGGAGGGAGC | CACCAGGCCC | GGCCTCCAGA | GGCAGGAGAG | TGTGTTCTGA | 1080 |

(31) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 359 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Met Gln Val Pro Asn Ser Thr Gly Pro Asp Asn Ala Thr Leu Gln Met 1 5 10 15

Leu Arg Asn Pro Ala Ile Ala Val Ala Leu Pro Val Val Tyr Ser Leu 20 25 30

30 Val Ala Ala Val Ser Ile Pro Gly Asn Leu Phe Ser Leu Trp Val Leu

| | | | 35 | | | | | 40 | | | | | 45 | | | |
|----|------------|------------|--------------|------------|------------|------------|--------------|------------|------------|------------|------------|------------|--------------|------------|------------|------------|
| | Cys | Arg 50 | Arg | Met | Gly | Pro | Arg 55 | Ser | Pro | Ser | Val | Ile 60 | Phe | Met | Ile | Asn |
| 5 | Leu 65 | Ser | Val | Thr | Asp | Leu 70 | Met | Leu | Ala | Ser | Val 75 | Leu | Pro | Phe | Gln | Ile 80 |
| | Tyr | Tyr | His | Cys | Asn 85 | Arg | His | His | Trp | Val 90 | Phe | Gly | Val | Leu | Leu 95 | Cys |
| | Asn | Val | Val | Thr 100 | Val | Ala | Phe | Tyr | Ala 105 | Asn | Met | Tyr | Ser | Ser 110 | Ile | Leu |
| 10 | Thr | Met | Thr 115 | Cys | Ile | Ser | Val | Glu 120 | Arg | Phe | Leu | Gly | Val 125 | Leu | Tyr | Pro |
| | Leu | Ser 130 | Ser | Lys | Arg | Trp | Arg 135 | Arg | Arg | Arg | Tyr | Ala 140 | Val | Ala | Ala | Cys |
| 15 | Ala 145 | Gly | Thr | Trp | Leu | Leu 150 | Leu | Leu | Thr | Ala | Leu 155 | Cys | Pro | Leu | Ala | Arg 160 |
| | Thr | Asp | Leu | Thr | Tyr 165 | Pro | Val | His | Ala | Leu 170 | Gly | Ile | Ile | Thr | Cys 175 | Phe |
| | Asp | Val | Leu | Lys 180 | | Thr | Met | Leu | Pro 185 | | Val | Ala | Met | Trp 190 | Ala | Val |
| 20 | Phe | Leu | Phe 195 | | Ile | Phe | Ile | Leu 200 | Leu | Phe | Leu | Ile | Pro 205 | | Val | Ile |
| | Thr | Val 210 | | Cys | Tyr | Thr | Ala 215 | | Ile | Leu | Lys | Leu 220 | | Arg | Thr | Glu |
| 25 | Glu 225 | | His | Gly | Arg | Glu 230 | | Arg | Arg | Arg | Ala 235 | | . Gly | Leu | Ala | Ala 240 |
| | Val | . Val | L Leu | Leu | Ala 245 | | e Val | Thr | Cys | 250 | | . Pro | Asn | . Asn | Phe 255 | Val |
| | Leu | ı Leı | ı Ala | His 260 | | · Val | L Ser | Arg | Leu 265 | | Yyr | : Gly | / Lys | Ser 270 | Tyr | Tyr |
| 30 | His | val | l Tyr 275 | | s Lei | ı Thi | r Lev | 280 | | ı Ser | c Cys | Let | 1 Asr 285 | | ı Cys | . Leu |
| | Ası | 290 | | e Val | l Tyi | ту: | r Phe 295 | | a Sei | r Arg | g Glı | 30 | | ı Lev | ı Arg | J Leu |
| | Arg | g Gl | u Ty | r Lei | ı Gly | у Су: | s Arg | g Arg | y Vai | l Pro | o Arg | J As | p Th | r Lei | ı Ası | Thr |

315

330 335

Arg Arg Glu Ser Leu Phe Ser Ala Arg Thr Thr Ser Val Arg Ser Glu

15

20

25

Ala Gly Ala His Pro Glu Gly Met Glu Gly Ala Thr Arg Pro Gly Leu 340 345 350

Gln Arg Gln Glu Ser Val Phe 355

- 5 (32) INFORMATION FOR SEQ ID NO:31:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1503 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
- 10 (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

| | ATGGAGCGTC | CCTGGGAGGA | CAGCCCAGGC | CCGGAGGGGG | CAGCTGAGGG | CTCGCCTGTG | 60 |
|---|------------|------------|------------|--------------|------------|------------|-------|
| | CCAGTCGCCG | ccggggcgcg | CTCCGGTGCC | GCGGCGAGTG | GCACAGGCTG | GCAGCCATGG | . 120 |
| | GCTGAGTGCC | CGGGACCCAA | GGGGAGGGG | CAACTGCTGG | CGACCGCCGG | CCCTTTGCGT | 180 |
| | CGCTGGCCCG | CCCCCTCGCC | TGCCAGCTCC | AGCCCCGCCC | CCGGAGCGGC | GTCCGCTCAC | 240 |
| | TCGGTTCAAG | GCAGCGCGAC | TGCGGGTGGC | GCACGACCAG | GGCGCAGACC | TTGGGGCGCG | 300 |
| | CGGCCCATGG | AGTCGGGGCT | GCTGCGGCCG | GCGCCGGTGA | GCGAGGTCAT | CGTCCTGCAT | 360 |
| | TACAACTACA | CCGGCAAGCT | CCGCGGTGCG | AGCTACCAGC | CGGGTGCCGG | CCTGCGCGCC | 420 |
| | GACGCCGTGG | TGTGCCTGGC | GGTGTGCGCC | TTCATCGTGC | TAGAGAATCT | AGCCGTGTTG | 480 |
| | TTGGTGCTCG | GACGCCACCC | GCGCTTCCAC | GCTCCCATGT | TCCTGCTCCT | GGGCAGCCTC | 540 |
| | ACGTTGTCGG | ATCTGCTGGC | AGGCGCCGCC | TACGCCGCCA | ACATCCTACT | GTCGGGGCCG | 600 |
| | CTCACGCTGA | AACTGTCCCC | CGCGCTCTGG | TTCGCACGGG | AGGGAGGCGT | CTTCGTGGCA | 660 |
| | CTCACTGCGT | CCGTGCTGAG | CCTCCTGGCC | ATCGCGCTGG | AGCGCAGCCT | CACCATGGCG | 720 |
| | CGCAGGGGGC | CCGCGCCCGT | CTCCAGTCGG | GGGCGCACGC | TGGCGATGGC | AGCCGCGGCC | 780 |
| | TGGGGCGTGT | CGCTGCTCCT | CGGGCTCCTG | CCAGCGCTGG | GCTGGAATTG | CCTGGGTCGC | 840 |
| | CTGGACGCTT | GCTCCACTGT | CTTGCCGCTC | TACGCCAAGG | CCTACGTGCT | CTTCTGCGTG | 900 |
| | CTCGCCTTCG | TGGGCATCCT | GGCCGCGATC | TGTGCACTCT | ACGCGCGCAT | CTACTGCCAG | 960 |
| | GTACGCGCCA | ACGCGCGGCG | CCTGCCGGCA | . CGGCCCGGGA | CTGCGGGGAC | CACCTCGACC | 1020 |
| ı | CGGGCGCGTC | GCAAGCCGCG | CTCTCTGGCC | TTGCTGCGCA | CGCTCAGCGT | GGTGCTCCTG | 1080 |

| | === | |
|----------|------------------|-----|
| 7 | | 5 |
| | Ī | |
| ÷ | | 200 |
| Junit. | Strate Strate | |
| man | al a | |
| i i | free, | |
| ww. | Herek | |
| ä | | |
| ·mir | === === | |
| | - | - |
| THEFT. | # | |
| ċ | : | |
| - | 222 | |
| 12111111 | 227 | |

| GCCTTTGTGG | CATGTTGGGG | CCCCCTCTTC | CTGCTGCTGT | TGCTCGACGT | GGCGTGCCCG | 1140 |
|------------|------------|------------|------------|------------|------------|------|
| GCGCGCACCT | GTCCTGTACT | CCTGCAGGCC | GATCCCTTCC | TGGGACTGGC | CATGGCCAAC | 1200 |
| TCACTTCTGA | ACCCCATCAT | CTACACGCTC | ACCAACCGCG | ACCTGCGCCA | CGCGCTCCTG | 1260 |
| CGCCTGGTCT | GCTGCGGACG | CCACTCCTGC | GGCAGAGACC | CGAGTGGCTC | CCAGCAGTCG | 1320 |
| GCGAGCGCGG | CTGAGGCTTC | CGGGGGCCTG | CGCCGCTGCC | TGCCCCCGGG | CCTTGATGGG | 1380 |
| AGCTTCAGCG | GCTCGGAGCG | CTCATCGCCC | CAGCGCGACG | GGCTGGACAC | CAGCGGCTCC | 1440 |
| ACAGGCAGCC | CCGGTGCACC | CACAGCCGCC | CGGACTCTGG | TATCAGAACC | GGCTGCAGAC | 1500 |
| TGA | | | | | | 1503 |

- (33) INFORMATION FOR SEQ ID NO:32:
- 10 (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 500 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- 15 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Met Glu Arg Pro Trp Glu Asp Ser Pro Gly Pro Glu Gly Ala Ala Glu
1 5 10 15

Gly Ser Pro Val Pro Val Ala Ala Gly Ala Arg Ser Gly Ala Ala Ala 20 25 30

Ser Gly Thr Gly Trp Gln Pro Trp Ala Glu Cys Pro Gly Pro Lys Gly 35 40 45

Arg Gly Gln Leu Leu Ala Thr Ala Gly Pro Leu Arg Arg Trp Pro Ala 50 55 60

Pro Ser Pro Ala Ser Ser Pro Ala Pro Gly Ala Ala Ser Ala His
70 75 80

Ser Val Gln Gly Ser Ala Thr Ala Gly Gly Ala Arg Pro Gly Arg Arg 85 90 95

Pro Trp Gly Ala Arg Pro Met Glu Ser Gly Leu Leu Arg Pro Ala Pro 30 100 105 110

Val Ser Glu Val Ile Val Leu His Tyr Asn Tyr Thr Gly Lys Leu Arg 115 120 125

Gly Ala Ser Tyr Gln Pro Gly Ala Gly Leu Arg Ala Asp Ala Val Val

: I

H

mum mum

113

And the Coll will will

445 435 440 Gly Leu Arg Arg Cys Leu Pro Pro Gly Leu Asp Gly Ser Phe Ser Gly 450 Ser Glu Arg Ser Ser Pro Gln Arg Asp Gly Leu Asp Thr Ser Gly Ser 5

Thr Gly Ser Pro Gly Ala Pro Thr Ala Ala Arg Thr Leu Val Ser Glu

Pro Ala Ala Asp 500

(34) INFORMATION FOR SEQ ID NO:33: 10

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1029 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear 15
 - (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ATGCAAGCCG TCGACAATCT CACCTCTGCG CCTGGGAACA CCAGTCTGTG CACCAGAGAC TACAAAATCA CCCAGGTCCT CTTCCCACTG CTCTACACTG TCCTGTTTTT TGTTGGACTT 120 20 ATCACAAATG GCCTGGCGAT GAGGATTTTC TTTCAAATCC GGAGTAAATC AAACTTTATT 180 ATTTTTCTTA AGAACACAGT CATTTCTGAT CTTCTCATGA TTCTGACTTT TCCATTCAAA 240 ATTCTTAGTG ATGCCAAACT GGGAACAGGA CCACTGAGAA CTTTTGTGTG TCAAGTTACC 300 TCCGTCATAT TTTATTTCAC AATGTATATC AGTATTTCAT TCCTGGGACT GATAACTATC 360 GATCGCTACC AGAAGACCAC CAGGCCATTT AAAACATCCA ACCCCAAAAA TCTCTTGGGG 420 25 GCTAAGATTC TCTCTGTTGT CATCTGGGCA TTCATGTTCT TACTCTCTTT GCCTAACATG 480 ATTCTGACCA ACAGGCAGCC GAGAGACAAG AATGTGAAGA AATGCTCTTT CCTTAAATCA 540 GAGTTCGGTC TAGTCTGGCA TGAAATAGTA AATTACATCT GTCAAGTCAT TTTCTGGATT 600 AATTTCTTAA TTGTTATTGT ATGTTATACA CTCATTACAA AAGAACTGTA CCGGTCATAC 660 GTAAGAACGA GGGGTGTAGG TAAAGTCCCC AGGAAAAAGG TGAACGTCAA AGTTTTCATT 720 30 ATCATTGCTG TATTCTTTAT TTGTTTTGTT CCTTTCCATT TTGCCCGAAT TCCTTACACC 780 CTGAGCCAAA CCCGGGATGT CTTTGACTGC ACTGCTGAAA ATACTCTGTT CTATGTGAAA 840

| ģ | Ö |
|----------------|--------------|
| 31111 | |
| 5 | |
| ·min. | Ţ |
| man | |
| | 9 |
| 2331337 | 120 |
| 7 | |
| mus. | 111) 111) |
| | Į.į |
| Water Total | 3 |
| ÷ | - |
| 1 | 2251 2267 |
| 114444 | :á: |

GAGAGCACTC TGTGGTTAAC TTCCTTAAAT GCATGCCTGG ATCCGTTCAT CTATTTTTC 900

CTTTGCAAGT CCTTCAGAAA TTCCTTGATA AGTATGCTGA AGTGCCCCAA TTCTGCAACA 960

TCTCTGTCCC AGGACAATAG GAAAAAAGAA CAGGATGGTG GTGACCCAAA TGAAGAGACT 1020

CCAATGTAA 1029

- 5 (35) INFORMATION FOR SEQ ID NO:34:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 342 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
- 10 (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Met Gln Ala Val Asp Asn Leu Thr Ser Ala Pro Gly Asn Thr Ser Leu 1 5 10 15

Cys Thr Arg Asp Tyr Lys Ile Thr Gln Val Leu Phe Pro Leu Leu Tyr 20 25 30

Thr Val Leu Phe Phe Val Gly Leu Ile Thr Asn Gly Leu Ala Met Arg 35 40 45

Ile Phe Phe Gln Ile Arg Ser Lys Ser Asn Phe Ile Ile Phe Leu Lys 50 55 60

Asn Thr Val Ile Ser Asp Leu Leu Met Ile Leu Thr Phe Pro Phe Lys 65 70 75 80

Ile Leu Ser Asp Ala Lys Leu Gly Thr Gly Pro Leu Arg Thr Phe Val 85 90 95

25 Cys Gln Val Thr Ser Val Ile Phe Tyr Phe Thr Met Tyr Ile Ser Ile 100 105 110

> Ser Phe Leu Gly Leu Ile Thr Ile Asp Arg Tyr Gln Lys Thr Thr Arg 115 120 125

Pro Phe Lys Thr Ser Asn Pro Lys Asn Leu Leu Gly Ala Lys Ile Leu 30 130 135 140

Ser Val Val Ile Trp Ala Phe Met Phe Leu Leu Ser Leu Pro Asn Met 145 150 155 160

Ile Leu Thr Asn Arg Gln Pro Arg Asp Lys Asn Val Lys Lys Cys Ser 165 170 175

35 Phe Leu Lys Ser Glu Phe Gly Leu Val Trp His Glu Ile Val Asn Tyr

| .min. | ### ### |
|------------|----------------|
| ÷ | ū |
| .13311. | T |
| ÷ | r. |
| THE P | 3 5 |
| tion: | Ţ, |
| THE PERSON | Hard Person |
| 2111141 | 1 |
| Ē | |
| Time: | === |
| THE PERSON | Ţ |
| -1111r | 20.j |
| ÷ | 녆 |
| ann. | |
| thint. | cắz |

| | | | | | 180 | | | | | 185 | | | | | 190 | | | |
|----|--|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-----|
| | | Ile | Cys | Gln 195 | Val | Ile | Phe | Trp | Ile 200 | Asn | Phe | Leu | Ile | Val 205 | Ile | Val | Cys | |
| 5 | | Tyr | Thr 210 | Leu | Ile | Thr | Lys | Glu 215 | Leu | Tyr | Arg | Ser | Tyr 220 | Val | Arg | Thr | Arg | |
| | | Gly 225 | Val | Gly | Lys | Val | Pro 230 | Arg | Lys | Lys | Val | Asn 235 | Val | Lys | Val | Phe | Ile 240 | |
| | | Ile | Ile | Ala | Val | Phe 245 | Phe | Ile | Cys | Phe | Val 250 | Pro | Phe | His | Phe | Ala 255 | Arg | |
| 10 | | Ile | Pro | Tyr | Thr 260 | Leu | Ser | Gln | Thr | Arg 265 | Asp | Val | Phe | Asp | Cys 270 | Thr | Ala | |
| | | Glu | Asn | Thr 275 | Leu | Phe | Tyr | Val | Lys 280 | Glu | Ser | Thr | Leu | Trp 285 | Leu | Thr | Ser | |
| 15 | | Leu | Asn 290 | Ala | Cys | Leu | Asp | Pro 295 | Phe | Ile | Tyr | Phe | Phe 300 | Leu | Cys | Lys | Ser : | |
| | | Phe 305 | Arg | Asn | Ser | Leu | Ile 310 | Ser | Met | Leu | Lys | Cys 315 | Pro | Asn | Ser | Ala | Thr 320 | |
| | | Ser | Leu | Ser | Gln | Asp 325 | Asn | Arg | Lys | Lys | Glu 330 | | Asp | Gly | Gly | Asp 335 | Pro | |
| 20 | | Asn | Glu | Glu | Thr 340 | Pro | Met | | | | | | | | | | | |
| | (36) | INF | ORMA | TION | FOR | SEQ | ID | NO:3 | 5: | | | | | | | | | |
| 25 | (36) INFORMATION FOR SEQ ID NO:35: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1077 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | | | | | | | | | | | | | | | | | |
| | | (ii) | MOL | ECUL | E TY | PE: | DNA | (gen | omic | :) | | | | | | | | |
| | | (xi) | SEQ | UENC | E DE | SCRI | PTIC | N: S | EQ I | D NC | :35: | | | | | | | |
| 30 | ATGT | 'CGGT | CT G | CTAC | CGTC | c cc | CAGG | GAAC | GAG | BACAC | TGC | TGAG | CTG | AA G | ACTT | CGCG | G | 60 |
| | GCCA | CAGG | CA C | AGCC | TTCC | T GC | TGCT | rggce | GCG | CTGC | TGG | GGCI | GCCI | GG C | CAACG | GCTT | .C | 120 |
| | GTGG | TGTG | GA G | CTTG | GCGG | G CI | 'GGCG | GCCI | GC | ACGGG | GGC | GACC | GCT | GC G | GCCI | CGCI | T | 180 |

GTGCTGCACC TGGCGCTGGC CGACGGCGCG GTGCTGCTGC TCACGCCGCT CTTTGTGGCC

TTCCTGACCC GGCAGGCCTG GCCGCTGGGC CAGGCGGGCT GCAAGGCGGT GTACTACGTG

240

| | TGCGCGCTCA | GCATGTACGC | CAGCGTGCTG | CTCACCGGCC | TGCTCAGCCT | GCAGCGCTGC | 360 |
|----|------------|------------|------------|------------|------------|------------|------|
| | CTCGCAGTCA | CCCGCCCCTT | CCTGGCGCCT | CGGCTGCGCA | GCCCGGCCCT | GGCCCGCCGC | 420 |
| | CTGCTGCTGG | CGGTCTGGCT | GGCCGCCCTG | TTGCTCGCCG | TCCCGGCCGC | CGTCTACCGC | 480 |
| | CACCTGTGGA | GGGACCGCGT | ATGCCAGCTG | TGCCACCCGT | CGCCGGTCCA | CGCCGCCGCC | 540 |
| 5 | CACCTGAGCC | TGGAGACTCT | GACCGCTTTC | GTGCTTCCTT | TCGGGCTGAT | GCTCGGCTGC | 600 |
| | TACAGCGTGA | CGCTGGCACG | GCTGCGGGGC | GCCCGCTGGG | GCTCCGGGCG | GCACGGGGCG | 660 |
| | CGGGTGGGCC | GGCTGGTGAG | CGCCATCGTG | CTTGCCTTCG | GCTTGCTCTG | GGCCCCCTAC | 720 |
| | CACGCAGTCA | ACCTTCTGCA | GGCGGTCGCA | GCGCTGGCTC | CACCGGAAGG | GGCCTTGGCG | 780 |
| | AAGCTGGGCG | GAGCCGGCCA | GGCGGCGCGA | GCGGGAACTA | CGGCCTTGGC | CTTCTTCAGT | 840 |
| 10 | TCTAGCGTCA | ACCCGGTGCT | CTACGTCTTC | ACCGCTGGAG | ATCTGCTGCC | CCGGGCAGGT | 900 |
| | CCCCGTTTCC | TCACGCGGCT | CTTCGAAGGC | TCTGGGGAGG | CCCGAGGGGG | CGGCCGCTCT | 960 |
| | AGGGAAGGGA | CCATGGAGCT | CCGAACTACC | CCTCAGCTGA | AAGTGGTGGG | GCAGGGCCGC | 1020 |
| | GGCAATGGAG | ACCCGGGGGG | TGGGATGGAG | AAGGACGGTC | CGGAATGGGA | CCTTTGA | 1077 |

- (37) INFORMATION FOR SEQ ID NO:36:
- 15 (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 358 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- 20 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Met Ser Val Cys Tyr Arg Pro Pro Gly Asn Glu Thr Leu Leu Ser Trp 1 5 10 15

Lys Thr Ser Arg Ala Thr Gly Thr Ala Phe Leu Leu Ala Ala Leu 25 20 25 30

Leu Gly Leu Pro Gly Asn Gly Phe Val Val Trp Ser Leu Ala Gly Trp

Arg Pro Ala Arg Gly Arg Pro Leu Ala Ala Thr Leu Val Leu His Leu 50 55 60

Ala Leu Ala Asp Gly Ala Val Leu Leu Thr Pro Leu Phe Val Ala 65 70 75 80

Phe Leu Thr Arg Gln Ala Trp Pro Leu Gly Gln Ala Gly Cys Lys Ala

PATENT

| | | | | | 85 | | | | | 90 | | | | | 95 | |
|----|------------|------------|------------|------------|------------|------------|------------|------------|--------------|------------|------------|------------|------------|------------|------------|------------|
| | Val | Tyr | Tyr | Val 100 | Cys | Ala | Leu | Ser | Met 105 | Tyr | Ala | Ser | Val | Leu 110 | Leu | Thr |
| 5 | Gly | Leu | Leu 115 | Ser | Leu | Gln | Arg | Cys 120 | Leu | Ala | Val | Thr | Arg 125 | Pro | Phe | Leu |
| | Ala | Pro 130 | Arg | Leu | Arg | Ser | Pro 135 | Ala | Leu | Ala | Arg | Arg 140 | Leu | Leu | Leu | Ala |
| | Val 145 | Trp | Leu | Ala | Ala | Leu 150 | Leu | Leu | Ala | Val | Pro 155 | Ala | Ala | Val | Tyr | Arg 160 |
| 10 | His | Leu | Trp | Arg | Asp 165 | Arg | Val | Cys | Gln | Leu 170 | Cys | His | Pro | Ser | Pro 175 | Val |
| | His | Ala | Ala | Ala 180 | His | Leu | Ser | Leu | Glu 185 | Thr | Leu | Thr | Ala | Phe 190 | Val | Leu |
| 15 | Pro | Phe | Gly 195 | Leu | Met | Leu | Gly | Cys 200 | Tyr | Ser | Val | Thr | Leu 205 | Ala | Arg | Leu |
| | Arg | Gly 210 | Ala | Arg | Trp | Gly | Ser 215 | Gly | Arg | His | Gly | Ala 220 | Arg | Val | Gly | Arg |
| | Leu 225 | Val | Ser | Ala | Ile | Val 230 | Leu | Ala | Phe | Gly | Leu 235 | Leu | Trp | Ala | Pro | Tyr 240 |
| 20 | His | Ala | Val | Asn | Leu 245 | Leu | Gln | Ala | Val | Ala 250 | Ala | Leu | Ala | Pro | Pro 255 | Glu |
| | Gly | Ala | Leu | Ala 260 | Lys | Leu | Gly | Gly | Ala 265 | | Gln | Ala | Ala | Arg 270 | Ala | Gly |
| 25 | Thr | Thr | Ala 275 | Leu | Ala | Phe | Phe | Ser 280 | | Ser | Val | Asn | Pro 285 | | Leu | Tyr |
| | Val | Phe 290 | | Ala | Gly | Asp | | Leu | | | Ala | Gly 300 | | Arg | Phe | Leu |
| | Thr 305 | _ | Leu | Phe | Glu | Gly 310 | | Gly | Glu | Ala | Arg 315 | | Gly | Gly | Arg | Ser 320 |
| 30 | Arg | Glu | Gly | Thr | Met 325 | | . Leu | Arg | Thr | Thr 330 | | Gln | Leu | . Lys | Val 335 | Val |
| | Gly | Gln | Gly | Arg | | Asn | Gly | Asp |) Pro 345 | | Gly | Gly | Met | 350 | | Asp |
| 35 | Gly | Pro | Glu 355 | _ | Asp | Leu | ι | | | | | | | | | |
| | | | | | | | | | | | | | | | | |

(38) INFORMATION FOR SEQ ID NO:37:

15

20

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1005 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
- 5 (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

| ATGCTGGGGA | rcatggcatg | GAATGCAACT | TGCAAAAACT | GGCTGGCAGC | AGAGGCTGCC | 60 |
|--------------|------------|------------|------------|------------|------------|-------|
| CTGGAAAAGT A | ACTACCTTTC | CATTTTTTAT | GGGATTGAGT | TCGTTGTGGG | AGTCCTTGGA | 120 |
| AATACCATTG | TTGTTTACGG | CTACATCTTC | TCTCTGAAGA | ACTGGAACAG | CAGTAATATT | 180 |
| TATCTCTTTA | ACCTCTCTGT | CTCTGACTTA | GCTTTTCTGT | GCACCCTCCC | CATGCTGATA | 240 |
| AGGAGTTATG | CCAATGGAAA | CTGGATATAT | GGAGACGTGC | TCTGCATAAG | CAACCGATAT | 300 |
| GTGCTTCATG | CCAACCTCTA | TACCAGCATT | CTCTTTCTCA | CTTTTATCAG | CATAGATCGA | . 360 |
| TACTTGATAA | TTAAGTATCC | TTTCCGAGAA | CACCTTCTGC | AAAAGAAAGA | GTTTGCTATT | 420 |
| TTAATCTCCT | TGGCCATTTG | GGTTTTAGTA | ACCTTAGAGT | TACTACCCAT | ACTTCCCCTT | 480 |
| ATAAATCCTG | TTATAACTGA | CAATGGCACC | ACCTGTAATG | ATTTTGCAAG | TTCTGGAGAC | 540 |
| CCCAACTACA | ACCTCATTTA | CAGCATGTGT | CTAACACTGT | TGGGGTTCCT | TATTCCTCTT | 600 |
| TTTGTGATGT | GTTTCTTTTA | TTACAAGATT | GCTCTCTTCC | TAAAGCAGAG | GAATAGGCAG | 660 |
| GTTGCTACTG | CTCTGCCCCT | TGAAAAGCCT | CTCAACTTGG | TCATCATGGC | AGTGGTAATC | 720 |
| TTCTCTGTGC | TTTTTACACC | CTATCACGTC | ATGCGGAATG | TGAGGATCGC | TTCACGCCTG | 780 |
| GGGAGTTGGA | AGCAGTATCA | GTGCACTCAG | GTCGTCATCA | ACTCCTTTTA | CATTGTGACA | 840 |
| CGGCCTTTGG | CCTTTCTGAA | CAGTGTCATC | AACCCTGTCT | TCTATTTTCT | TTTGGGAGAT | 900 |
| CACTTCAGGG | ACATGCTGAT | GAATCAACTG | AGACACAACT | TCAAATCCCT | TACATCCTTT | 960 |
| AGCAGATGGG | CTCATGAACT | CCTACTTTCA | TTCAGAGAAA | AGTGA | | 1005 |

- 25 (39) INFORMATION FOR SEQ ID NO:38:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 334 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein

| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38: | | | | | | | | | | | | | | | |
|----|--|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|--------------|------------|
| | Met 1 | Leu | Gly | Ile | Met 5 | Ala | Trp | Asn | Ala | Thr 10 | Cys | Lys | Asn | Trp | Leu 15 | Ala |
| 5 | Ala | Glu | Ala | Ala 20 | Leu | Glu | Lys | Tyr | Tyr 25 | Leu | Ser | Ile | Phe | Tyr 30 | Gly | Ile |
| | Glu | Phe | Val 35 | Val | Gly | Val | Leu | Gly 40 | Asn | Thr | Ile | Val | Val 45 | Tyr | Gly | Tyr |
| | Ile | Phe 50 | Ser | Leu | Lys | Asn | Trp 55 | Asn | Ser | Ser | Asn | Ile 60 | Tyr | Leu | Phe | Asn |
| 10 | Leu 65 | Ser | Val | Ser | Asp | Leu 70 | Ala | Phe | Leu | Cys | Thr 75 | Leu | Pro | Met | Leu | Ile 80 |
| | Arg | Ser | Tyr | Ala | Asn 85 | Gly | Asn | Trp | Ile | Tyr 90 | Gly | Asp | Val | Leu | Cys 95 | Ile |
| 15 | Ser | Asn | Arg | Tyr 100 | Val | Leu | His | Ala | Asn 105 | Leu | Tyr | Thr | Ser | Ile 110 | Leu | Phe |
| | Leu | Thr | Phe 115 | Ile | Ser | Ile | Asp | Arg 120 | Tyr | Leu | Ile | Ile | Lys 125 | Tyr | Pro | Phe |
| | Arg | Glu 130 | His | Leu | Leu | Gln | Lys 135 | Lys | Glu | Phe | Ala | Ile 140 | Leu | Ile | Ser | Leu |
| 20 | Ala 145 | Ile | Trp | Val | Leu | Val 150 | Thr | Leu | Glu | Leu | Leu 155 | Pro | Ile | Leu | Pro | Leu 160 |
| | Ile | Asn | Pro | Val | Ile 165 | Thr | Asp | Asn | Gly | Thr 170 | Thr | Cys | Asn | Asp | Phe 175 | Ala |
| 25 | Ser | Ser | Gly | Asp 180 | Pro | Asn | Tyr | Asn | Leu 185 | | Tyr | Ser | Met | Cys 190 | Leu | Thr |
| | Leu | Leu | Gly 195 | | Leu | Ile | Pro | Leu 200 | | · Val | Met | Cys | Phe 205 | | Tyr | Tyr |
| | Lys | Ile 210 | | Leu | Phe | Leu | Lys 215 | | Arg | Asn | Arg | Gln 220 | | Ala | Thr | Ala |
| 30 | Leu 225 | | Leu | Glu | . Lys | Pro 230 | | Asn | Leu | ı Val | 11e 235 | | Ala | Val | Val | 1le 240 |
| | Phe | Ser | Val | Leu | Phe 245 | | Pro | Tyr | His | Val 250 | | . Arg | Asr | ı Val | . Arg 255 | , Ile |
| 35 | Ala | . Ser | Arg | Leu 260 | | Ser | Trp | Lys | Glr 265 | | Glr | Cys | : Thr | Glr 270 | | Val |
| | Ile | Asn | ser | Phe | . Tyr | Ile | val | Thr | Arg | g Pro | Leu | ı Ala | n Phe | e Lev | ı Asr | ser |

10

| | | 275 | | | | | 280 | | | | | 285 | | | |
|-----|------------|-----|-----|-----|-----|------------|-----|-----|-----|-----|------------|-----|-----|-----|-----|
| Val | Ile 290 | Asn | Pro | Val | Phe | Tyr 295 | Phe | Leu | Leu | Gly | Asp 300 | His | Phe | Arg | Asp |
| Met | Leu | Met | Asn | Gln | Leu | Arq | His | Asn | Phe | Lys | Ser | Leu | Thr | Ser | Phe |

Ser Arg Trp Ala His Glu Leu Leu Ser Phe Arg Glu Lys

(40) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1296 base pairs

325

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

ATGCAGGCGC TTAACATTAC CCCGGAGCAG TTCTCTCGGC TGCTGCGGGA CCACAACCTG 60 ACGCGGGAGC AGTTCATCGC TCTGTACCGG CTGCGACCGC TCGTCTACAC CCCAGAGCTG 120 CCGGGACGCG CCAAGCTGGC CCTCGTGCTC ACCGGCGTGC TCATCTTCGC CCTGGCGCTC 180 TTTGGCAATG CTCTGGTGTT CTACGTGGTG ACCCGCAGCA AGGCCATGCG CACCGTCACC 240 20 AACATCTTTA TCTGCTCCTT GGCGCTCAGT GACCTGCTCA TCACCTTCTT CTGCATTCCC 300 GTCACCATGC TCCAGAACAT TTCCGACAAC TGGCTGGGGG GTGCTTTCAT TTGCAAGATG 360 GTGCCATTTG TCCAGTCTAC CGCTGTTGTG ACAGAAATGC TCACTATGAC CTGCATTGCT 420 GTGGAAAGGC ACCAGGGACT TGTGCATCCT TTTAAAATGA AGTGGCAATA CACCAACCGA 480 AGGGCTTTCA CAATGCTAGG TGTGGTCTGG CTGGTGGCAG TCATCGTAGG ATCACCCATG 540 ~ 25 TGGCACGTGC AACAACTTGA GATCAAATAT GACTTCCTAT ATGAAAAGGA ACACATCTGC 600 TGCTTAGAAG AGTGGACCAG CCCTGTGCAC CAGAAGATCT ACACCACCTT CATCCTTGTC 660 ATCCTCTTCC TCCTGCCTCT TATGGTGATG CTTATTCTGT ACAGTAAAAT TGGTTATGAA 720 CTTTGGATAA AGAAAAGAGT TGGGGATGGT TCAGTGCTTC GAACTATTCA TGGAAAAGAA 780 ATGTCCAAAA TAGCCAGGAA GAAGAAACGA GCTGTCATTA TGATGGTGAC AGTGGTGGCT 840 30 CTCTTTGCTG TGTGCTGGGC ACCATTCCAT GTTGTCCATA TGATGATTGA ATACAGTAAT 900 960 TTTGAAAAGG AATATGATGA TGTCACAATC AAGATGATTT TTGCTATCGT GCAAATTATT

| | | | _ | | | | |
|---|------------|--------------|-------------|------------|------------|------------|------|
| | GGATTTTCCA | ACTCCATCTG | TAATCCCATT | GTCTATGCAT | TTATGAATGA | AAACTTCAAA | 1020 |
| | AAAAATGTTT | TGTCTGCAGT | TTGTTATTGC | ATAGTAAATA | AAACCTTCTC | TCCAGCACAA | 1080 |
| | AGGCATGGAA | ATTCAGGAAT | TACAATGATG | CGGAAGAAAG | CAAAGTTTTC | CCTCAGAGAG | 1140 |
| | AATCCAGTGG | AGGAAACCAA | AGGAGAAGCA | TTCAGTGATG | GCAACATTGA | AGTCAAATTG | 1200 |
| 5 | TGTGAACAGA | CAGAGGAGAA | GAAAAAGCTC | AAACGACATC | TTGCTCTCTT | TAGGTCTGAA | 1260 |
| | CTGGCTGAGA | ATTCTCCTTT | AGACAGTGGG | CATTAA | | | 1296 |
| | (41) INFOR | MATION FOR | SEQ ID NO:4 | 0: | | | |
| | · - / | EQUENCE CHAI | | | | | |

- 10 (B) TYPE: amino acid (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:
- Met Gln Ala Leu Asn Ile Thr Pro Glu Gln Phe Ser Arg Leu Leu Arg

 1 10 15
 - Asp His Asn Leu Thr Arg Glu Gln Phe Ile Ala Leu Tyr Arg Leu Arg 20 25 30
- Pro Leu Val Tyr Thr Pro Glu Leu Pro Gly Arg Ala Lys Leu Ala Leu 20 35 40 45
 - Val Leu Thr Gly Val Leu Ile Phe Ala Leu Ala Leu Phe Gly Asn Ala 50 55 60
 - Leu Val Phe Tyr Val Val Thr Arg Ser Lys Ala Met Arg Thr Val Thr 65 70 75 80
- 25 Asn Ile Phe Ile Cys Ser Leu Ala Leu Ser Asp Leu Leu Ile Thr Phe 85 90 95
 - Phe Cys Ile Pro Val Thr Met Leu Gln Asn Ile Ser Asp Asn Trp Leu 100 105 110
- Gly Gly Ala Phe Ile Cys Lys Met Val Pro Phe Val Gln Ser Thr Ala 30 115 120 125
 - Val Val Thr Glu Met Leu Thr Met Thr Cys Ile Ala Val Glu Arg His 130 135 140
 - Gln Gly Leu Val His Pro Phe Lys Met Lys Trp Gln Tyr Thr Asn Arg 145 150 155 160

| | Arg | Ala | Phe | Thr | Met 165 | Leu | Gly | Val | Val | Trp 170 | Leu | Val | Ala | Val | Ile 175 | Val |
|----|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | Gly | Ser | Pro | Met 180 | Trp | His | Val | Gln | Gln 185 | Leu | Glu | Ile | Lys | Tyr 190 | Asp | Phe |
| 5 | Leu | Tyr | Glu 195 | Lys | Glu | His | Ile | Cys 200 | Cys | Leu | Glu | Glu | Trp 205 | Thr | Ser | Pro |
| | Val | His 210 | Gln | Lys | Ile | Tyr | Thr 215 | Thr | Phe | Ile | Leu | Val 220 | Ile | Leu | Phe | Leu |
| 10 | Leu 225 | Pro | Leu | Met | Val | Met 230 | Leu | Ile | Leu | Tyr | Ser 235 | Lys | Ile | Gly | Tyr | Glu 240 |
| | Leu | Trp | Ile | Lys | Lys 245 | Arg | Val | Gly | Asp | Gly 250 | Ser | Val | Leu | Arg | Thr 255 | Ile |
| | His | Gly | Lys | Glu 260 | Met | Ser | Lys | Ile | Ala 265 | Arg | Lys | Lys | Lys | Arg 270 | Ala | Val |
| 15 | Ile | Met | Met 275 | Val | Thr | Val | Val | Ala 280 | Leu | Phe | Ala | Val | Cys 285 | Trp | Ala | Pro |
| | Phe | His 290 | Val | Val | His | Met | Met 295 | Ile | Glu | Tyr | Ser | Asn 300 | Phe | Glu | Lys | Glu |
| 20 | Tyr 305 | Asp | Asp | Val | Thr | Ile 310 | Lys | Met | Ile | Phe | Ala 315 | Ile | Val | Gln | Ile | Ile 320 |
| | Gly | Phe | Ser | Asn | Ser 325 | Ile | Cys | Asn | Pro | Ile 330 | Val | Tyr | Ala | Phe | Met 335 | Asn |
| | Glu | Asn | Phe | Lys 340 | - | Asn | Val | Leu | Ser 345 | Ala | Val | Cys | Tyr | Cys 350 | Ile | Val |
| 25 | Asn | Lys | Thr 355 | | Ser | Pro | Ala | Gln 360 | Arg | His | Gly | Asn | Ser 365 | | Ile | Thr |
| | Met | Met 370 | | Lys | Lys | Ala | Lys 375 | | Ser | Leu | . Arg | Glu 380 | Asn | Pro | Val | Glu |
| 30 | Glu 385 | | Lys | Gly | Glu | Ala 390 | | Ser | Asp | Gly | Asn 395 | | Glu | ı Val | Lys | Leu 400 |
| | Cys | Glu | Gln | Thr | Glu 405 | | . Lys | Lys | Lys | 410 | | Arg | , His | . Leu | Ala 415 | Leu |
| | Phe | Arg | , Ser | Glu 420 | | ı Ala | Glu | ı Asn | Ser 425 | r Pro |) Leu | ı Asp | Se1 | Gly 430 | | 3 |

- 35 (42) INFORMATION FOR SEQ ID NO:41:
 - (i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 24 base pairs

(D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CTGTGTACAG CAGTTCGCAG AGTG

24

PATENT

- (43) INFORMATION FOR SEQ ID NO:42:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:
- 15 GAGTGCCAGG CAGAGCAGGT AGAC

10

20

30

١.,

Ш M

113

24

- (44) INFORMATION FOR SEQ ID NO:43:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:
- 25 CCCGAATTCC TGCTTGCTCC CAGCTTGGCC C

- (45) INFORMATION FOR SEQ ID NO:44:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iv) ANTI-SENSE: YES

| AREN-0054 | | - 112 - P. | | | | | |
|-----------|-----------|---|---------|--|--|--|--|
| | (xi) | SEQUENCE DESCRIPTION: SEQ ID NO:44: | | | | | |
| | TGTGGATC | CT GCTGTCAAAG GTCCCATTCC GG | 32 | | | | |
| | (46) INFO | DRMATION FOR SEQ ID NO:45: | | | | | |
| 5 | (i) | SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | | | | | |
| | (ii) | MOLECULE TYPE: DNA (genomic) | | | | | |
| 10 | (iv) | ANTI-SENSE: NO | | | | | |
| | (xi) | SEQUENCE DESCRIPTION: SEQ ID NO:45: | | | | | |
| | TCACAATG | CT AGGTGTGGTC | 20 | | | | |
| | (47) INF | ORMATION FOR SEQ ID NO:46: | • | | | | |
| 15 | (i) | SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | | | | | |
| | (ii) | MOLECULE TYPE: DNA (genomic) | | | | | |
| 20 | (iv) | ANTI-SENSE: YES | | | | | |
| | (xi) | SEQUENCE DESCRIPTION: SEQ ID NO:46: | | | | | |
| | TGCATAGA | CA ATGGGATTAC AG | 22 | | | | |
| | (48) INF | ORMATION FOR SEQ ID NO:47: | | | | | |
| 25 | (i) | SEQUENCE CHARACTERISTICS: (A) LENGTH: 511 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | | | | | |
| | (ii) | MOLECULE TYPE: DNA (genomic) | | | | | |
| 30 | (xi) | SEQUENCE DESCRIPTION: SEQ ID NO:47: | | | | | |
| | TCACAATO | GCT AGGTGTGGTC TGGCTGGTGG CAGTCATCGT AGGATCACCC ATGTGGC | ACG 60 | | | | |
| | TGCAACA | ACT TGAGATCAAA TATGACTTCC TATATGAAAA GGAACACATC TGCTGCT | TAG 120 | | | | |

- 113 -

AAGAGTGGAC CAGCCCTGTG CACCAGAAGA TCTACACCAC CTTCATCCTT GTCATCCTCT

TCCTCCTGCC TCTTATGGTG ATGCTTATTC TGTACGTAAA ATTGGTTATG AACTTTGGAT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CTGCTTAGAA GAGTGGACCA G

(iv) ANTI-SENSE: NO

21

PATENT

180

240

300

360

420

480

511

- (50) INFORMATION FOR SEQ ID NO:49:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
- 25 (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CTGTGCACCA GAAGATCTAC AC

22

- (51) INFORMATION FOR SEQ ID NO:50:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

30

15

20

AREN-0054

15

The same

the test that

AREN-0054

| | (ii) MOLECULE TYPE: DNA (genomic) | |
|----|--|----|
| | (iv) ANTI-SENSE: YES | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:50: | |
| | CAAGGATGAA GGTGGTGTAG A | 21 |
| 5 | (52) INFORMATION FOR SEQ ID NO:51: | |
| 10 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: YES | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51: | • |
| | GTGTAGATCT TCTGGTGCAC AGG | 23 |
| 15 | (53) INFORMATION FOR SEQ ID NO:52: | |
| 20 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) | |
| | | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52: | |
| | GCAATGCAGG TCATAGTGAG C | 21 |
| | (54) INFORMATION FOR SEQ ID NO:53: | |
| 25 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| 30 | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iii) HYPOTHETICAL: YES | |
| | (iv) ANTI-SENSE: YES | |

- 114 -

PATENT

| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53: | |
|----|--|----|
| | TGGAGCATGG TGACGGGAAT GCAGAAG | 27 |
| | (55) INFORMATION FOR SEQ ID NO:54: | • |
| 5 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| 10 | (iv) ANTI-SENSE: YES | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54: | |
| | GTGATGAGCA GGTCACTGAG CGCCAAG | 27 |
| | | |
| | (56) INFORMATION FOR SEQ ID NO:55: | |
| 15 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| 20 | (iv) ANTI-SENSE: NO | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55: | |
| | GCAATGCAGG CGCTTAACAT TAC | 23 |
| | (57) INFORMATION FOR SEQ ID NO:56: | |
| 25 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| 30 | (iv) ANTI-SENSE: YES | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56: | |
| | TTGGGTTACA ATCTGAAGGG CA | 22 |

- 115 -

PATENT

AREN-0054

| ARE | N-0054 | - 116 - | PATENT |
|-----|-----------|--|--------|
| | (58) INFO | RMATION FOR SEQ ID NO:57: | |
| 5 | (i) | SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) | MOLECULE TYPE: DNA (genomic) | |
| | (iv) | ANTI-SENSE: NO | |
| | (xi) | SEQUENCE DESCRIPTION: SEQ ID NO:57: | |
| 10 | ACTCCGTG | TC CAGCAGGACT CTG | 23 |
| | (58) INF | ORMATION FOR SEQ ID NO:58: | |
| 15 | (i) | SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) | MOLECULE TYPE: DNA (genomic) | |
| | (iv) | ANTI-SENSE: YES | |
| | (xi) | SEQUENCE DESCRIPTION: SEQ ID NO:58: | |
| 20 | TGCGTGTT | CC TGGACCCTCA CGTG | 24 |
| | (58) INF | ORMATION FOR SEQ ID NO:59: | |
| 25 | (i) | SEQUENCE CHARACTERISTICS: (A) LENGTH: 29 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) | MOLECULE TYPE: DNA (genomic) | |
| | (iv) | ANTI-SENSE: NO | |
| | (xi) | SEQUENCE DESCRIPTION: SEQ ID NO:59: | |
| 30 | CAGGCCTT | TGG ATTTTAATGT CAGGGATGG | 29 |
| | (61) IN | FORMATION FOR SEQ ID NO:60: | |

(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 27 base pairs

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

CCTGATTCAT TTAGGTGAGA TTGAGAC

(64) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs

(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear

(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

AREN-0054

5

30

- 117 -

PATENT

27

27

H

was man

IL.

إِنْ "

 AREN-0054

26

26

CCCAAGCTTC CCCAGGTGTA TTTGAT

(3) INFORMATION FOR SEQ ID NO:63:

ATGATTCTCA ACTCTTCTAC TGAAGATGGT ATTAAAAGAA TCCAAGATGA TTGTCCCAAA 60 GCTGGAAGGC ATAATTACAT ATTTGTCATG ATTCCTACTT TATACAGTAT CATCTTTGTG 120 GTGGGAATAT TTGGAAACAG CTTGGTGGTG ATAGTCATTT ACTTTTATAT GAAGCTGAAG 180 ACTGTGGCCA GTGTTTTCT TTTGAATTTA GCACTGGCTG ACTTATGCTT TTTACTGACT 240 25 TTGCCACTAT GGGCTGTCTA CACAGCTATG GAATACCGCT GGCCCTTTGG CAATTACCTA 300 TGTAAGATTG CTTCAGCCAG CGTCAGTTTC AACCTGTACG CTAGTGTGTT TCTACTCACG 360 TGTCTCAGCA TTGATCGATA CCTGGCTATT GTTCACCCAA TGAAGTCCCG CCTTCGACGC 420 ACAATGCTTG TAGCCAAAGT CACCTGCATC ATCATTTGGC TGCTGGCAGG CTTGGCCAGT 480 TTGCCAGCTA TAATCCATCG AAATGTATTT TTCATTGAGA ACACCAATAT TACAGTTTGT 540 30 GCTTTCCATT ATGAGTCCCA AAATTCAACC CTTCCGATAG GGCTGGGCCT GACCAAAAAT 600

| ATACTGGGTT | TCCTGTTTCC | TTTTCTGATC | ATTCTTACAA | GTTATACTCT | TATTTGGAAG | 660 |
|------------|------------|------------|------------|------------|------------|------|
| GCCCTAAAGA | AGGCTTATGA | AATTCAGAAG | AACAAACCAA | GAAATGATGA | TATTTTTAAG | 720 |
| ATAATTATGG | CAATTGTGCT | TTTCTTTTTC | TTTTCCTGGA | TTCCCCACCA | AATATTCACT | 780 |
| TTTCTGGATG | TATTGATTCA | ACTAGGCATC | ATACGTGACT | GTAGAATTGC | AGATATTGTG | 840 |
| GACACGGCCA | TGCCTATCAC | CATTTGTATA | GCTTATTTTA | ACAATTGCCT | GAATCCTCTT | 900 |
| TTTTATGGCT | TTCTGGGGAA | AAAATTTAAA | AGATATTTTC | TCCAGCTTCT | AAAATATATT | 960 |
| CCCCCAAAAG | CCAAATCCCA | CTCAAACCTT | TCAACAAAAA | TGAGCACGCT | TTCCTACCGC | 1020 |
| СССТСАGATA | ATGTAAGCTC | ATCCACCAAG | AAGCCTGCAC | CATGTTTTGA | GGTTGAGTGA | 1080 |

(67) INFORMATION FOR SEQ ID NO:66:

- 10 (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 359 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- 15 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

Met Ile Leu Asn Ser Ser Thr Glu Asp Gly Ile Lys Arg Ile Gln Asp 1 5 10 15

Asp Cys Pro Lys Ala Gly Arg His Asn Tyr Ile Phe Val Met Ile Pro 20 25 30

Thr Leu Tyr Ser Ile Ile Phe Val Val Gly Ile Phe Gly Asn Ser Leu 35 40 45

Val Val Ile Val Ile Tyr Phe Tyr Met Lys Leu Lys Thr Val Ala Ser 50 55 60

Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Leu Cys Phe Leu Leu Thr
65 70 75 80

Leu Pro Leu Trp Ala Val Tyr Thr Ala Met Glu Tyr Arg Trp Pro Phe 85 90 95

Gly Asn Tyr Leu Cys Lys Ile Ala Ser Ala Ser Val Ser Phe Asn Leu 30 100 105 110

> Tyr Ala Ser Val Phe Leu Leu Thr Cys Leu Ser Ile Asp Arg Tyr Leu 115 120 125

> Ala Ile Val His Pro Met Lys Ser Arg Leu Arg Arg Thr Met Leu Val

130 135 140 Ala Lys Val Thr Cys Ile Ile Ile Trp Leu Leu Ala Gly Leu Ala Ser 145 Leu Pro Ala Ile Ile His Arg Asn Val Phe Phe Ile Glu Asn Thr Asn 5 170 Ile Thr Val Cys Ala Phe His Tyr Glu Ser Gln Asn Ser Thr Leu Pro 185 Ile Gly Leu Gly Leu Thr Lys Asn Ile Leu Gly Phe Leu Phe Pro Phe Leu Ile Ile Leu Thr Ser Tyr Thr Leu Ile Trp Lys Ala Leu Lys Lys 10 215 Ala Tyr Glu Ile Gln Lys Asn Lys Pro Arg Asn Asp Asp Ile Phe Lys 230 Ile Ile Met Ala Ile Val Leu Phe Phe Phe Phe Ser Trp Ile Pro His 250 15 Gln Ile Phe Thr Phe Leu Asp Val Leu Ile Gln Leu Gly Ile Ile Arg Asp Cys Arg Ile Ala Asp Ile Val Asp Thr Ala Met Pro Ile Thr Ile 275 Cys Ile Ala Tyr Phe Asn Asn Cys Leu Asn Pro Leu Phe Tyr Gly Phe 20 295 Leu Gly Lys Lys Phe Lys Arg Tyr Phe Leu Gln Leu Leu Lys Tyr Ile 305 Pro Pro Lys Ala Lys Ser His Ser Asn Leu Ser Thr Lys Met Ser Thr 25 325 Leu Ser Tyr Arg Pro Ser Asp Asn Val Ser Ser Ser Thr Lys Lys Pro 350 345 Ala Pro Cys Phe Glu Val Glu 355 30 (68) INFORMATION FOR SEQ ID NO:67: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- 35 (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)

| ARE | N-0054 | - 121 - | PATENT |
|-----|----------|---|--------|
| | (xi) | SEQUENCE DESCRIPTION: SEQ ID NO:67: | |
| | ACCATGGG | CA GCCCCTGGAA CGGCAGC | 27 |
| | (69) INF | ORMATION FOR SEQ ID NO:68: | |
| 5 | (i) | SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) | MOLECULE TYPE: DNA (genomic) | |
| 10 | (xi) | SEQUENCE DESCRIPTION: SEQ ID NO:68: | |
| | AGAACCAC | CA CCAGCAGGAC GCGGACGGTC TGCCGGTGG | 39 |
| | (70) INF | CORMATION FOR SEQ ID NO:69: | |
| 15 | (i) | SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) | MOLECULE TYPE: DNA (genomic) | |
| | (xi) | SEQUENCE DESCRIPTION: SEQ ID NO:69: | |
| 20 | GTCCGCG' | ICC TGCTGGTGGT GGTTCTGGCA TTTATAATT | 39 |
| | (71) IN | FORMATION FOR SEQ ID NO:70: | |
| 25 | (i |) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: not relevant | |
| | (ii |) MOLECULE TYPE: DNA (genomic) | |
| | (xi) SE | QUENCE DESCRIPTION: SEQ ID NO:70: | |
| | CCTGGAT | CCT TATCCCATCG TCTTCACGTT AGC | 33 |
| 30 | (72) IN | FORMATION FOR SEQ ID NO:71: | |
| | (i |) SEQUENCE CHARACTERISTICS: (A) LENGTH: 26 base pairs (B) TYPE: nucleic acid | |

(C) STRANDEDNESS: single

| CHEEK. | |
|--------|--------------|
| à | |
| illi. | |
| ÷ | 5 |
| ann. | Ħ |
| mun | il in it |
| | |
| 10000 | Į. |
| 3 | |
| | 2007 2007 |
| .min. | Ţ. |
| · | 227 <u>1</u> |
| 4 | Į. |
| | 2222 |
| man | 7Ž: |

| | (D) TOPOLOGY: linear | |
|----|---|-----|
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: NO | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71: | |
| 5 | CTGGAATTCT CCTGCCAGCA TGGTGA 26 | |
| | (73) INFORMATION FOR SEQ ID NO:72: | |
| 10 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: YES | • |
| 15 | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:72: | |
| | GCAGGATCCT ATATTGCGTG CTCTGTCCCC | |
| | (74) INFORMATION FOR SEQ ID NO:73: | |
| 20 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 999 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| 25 | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73: | |
| | ATGGTGAACT CCACCCACCG TGGGATGCAC ACTTCTCTGC ACCTCTGGAA CCGCAGCAGT | 60 |
| , | TACAGACTGC ACAGCAATGC CAGTGAGTCC CTTGGAAAAG GCTACTCTGA TGGAGGGTGC | 120 |
| | TACGAGCAAC TTTTTGTCTC TCCTGAGGTG TTTGTGACTC TGGGTGTCAT CAGCTTGTTG | 180 |
| | GAGAATATCT TAGTGATTGT GGCAATAGCC AAGAACAAGA ATCTGCATTC ACCCATGTAC | 240 |
| 30 | TTTTTCATCT GCAGCTTGGC TGTGGCTGAT ATGCTGGTGA GCGTTTCAAA TGGATCAGAA | 300 |

ACCATTATCA TCACCCTATT AAACAGTACA GATACGGATG CACAGAGTTT CACAGTGAAT

ATTGATAATG TCATTGACTC GGTGATCTGT AGCTCCTTGC TTGCATCCAT TTGCAGCCTG

360

PATENT

480

540

600

660

| CTTCACATTA | AGAGGATTGC | TGTCCTCCCC | GGCACTGGTG | CCATCCGCCA | AGGTGCCAAT | 720 |
|--------------------------|------------|------------|-------------|------------|------------|-----|
| ATGAAGGGAG | CGATTACCTT | GACCATCCTG | ATTGGCGTCT | TTGTTGTCTG | CTGGGCCCCA | 780 |
| TTCTTCCTCC | ACTTAATATT | CTACATCTCT | TGTCCTCAGA | ATCCATATTG | TGTGTGCTTC | 840 |
| ATGTCTCACT | TTAACTTGTA | TCTCATACTG | ATCATGTGTA | ATTCAATCAT | CGATCCTCTG | 900 |
| እ ምም ጥ እ ሞር ር አ ር | теседаетса | AGAACTGAGG | אאארכידייכא | AAGAGATCAT | CTGTTGCTAT | 960 |

- 123 -

10 CCCCTGGGAG GCCTTTGTGA CTTGTCTAGC AGATATTAA 999

(75) INFORMATION FOR SEQ ID NO:74:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 332 amino acids
 - (B) TYPE: amino acid
- 15 (C) STRANDEDNESS:

AREN-0054

- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

| | Met | Val | Asn | Ser | Thr | His | Arg | Gly | Met | His | Thr | Ser | Leu | His | Leu | Trp |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 20 | 1 | | | | 5 | | | | | 10 | | | | | 15 | |

Asn Arg Ser Ser Tyr Arg Leu His Ser Asn Ala Ser Glu Ser Leu Gly 20 25 30

Lys Gly Tyr Ser Asp Gly Gly Cys Tyr Glu Gln Leu Phe Val Ser Pro

25 Glu Val Phe Val Thr Leu Gly Val Ile Ser Leu Leu Glu Asn Ile Leu 50 55 60

Val Ile Val Ala Ile Ala Lys Asn Lys Asn Leu His Ser Pro Met Tyr 65 70 75 80

Phe Phe Ile Cys Ser Leu Ala Val Ala Asp Met Leu Val Ser Val Ser 85 90 95

Asn Gly Ser Glu Thr Ile Ile Ile Thr Leu Leu Asn Ser Thr Asp Thr 100 105 110

Asp Ala Gln Ser Phe Thr Val Asn Ile Asp Asn Val Ile Asp Ser Val

115 120 125 Ile Cys Ser Ser Leu Leu Ala Ser Ile Cys Ser Leu Leu Ser Ile Ala 135 Val Asp Arg Tyr Phe Thr Ile Phe Tyr Ala Leu Gln Tyr His Asn Ile 5 Met Thr Val Lys Arg Val Gly Ile Ser Ile Ser Cys Ile Trp Ala Ala 165 170 Cys Thr Val Ser Gly Ile Leu Phe Ile Ile Tyr Ser Asp Ser Ser Ala 185 Val Ile Ile Cys Leu Ile Thr Met Phe Thr Met Leu Ala Leu Met 10 200 Ala Ser Leu Tyr Val His Met Phe Leu Met Ala Arg Leu His Ile Lys 210 Arg Ile Ala Val Leu Pro Gly Thr Gly Ala Ile Arg Gln Gly Ala Asn 15 Met Lys Gly Ala Ile Thr Leu Thr Ile Leu Ile Gly Val Phe Val Val 245 250 Cys Trp Ala Pro Phe Phe Leu His Leu Ile Phe Tyr Ile Ser Cys Pro 265 Gln Asn Pro Tyr Cys Val Cys Phe Met Ser His Phe Asn Leu Tyr Leu 20 280 Ile Leu Ile Met Cys Asn Ser Ile Ile Asp Pro Leu Ile Tyr Ala Leu 290 295 Arg Ser Gln Glu Leu Arg Lys Thr Phe Lys Glu Ile Ile Cys Cys Tyr 25 310 Pro Leu Gly Gly Leu Cys Asp Leu Ser Ser Arg Tyr

- (76) INFORMATION FOR SEQ ID NO:75:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

20

25

30

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

GTGGAATTCA TTTGCCCTGC CTCAACCCCC A

31

10 (78) INFORMATION FOR SEQ ID NO:77:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1344 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
- 15 (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

| | ATGGAGCTGC | TAAAGCTGAA | CCGGAGCGTG | CAGGGAACCG | GACCCGGGCC | GGGGGCTTCC | 60 |
|---|------------|------------|------------|------------|------------|------------|-----|
| | CTGTGCCGCC | CGGGGGCGCC | TCTCCTCAAC | AGCAGCAGTG | TGGGCAACCT | CAGCTGCGAG | 120 |
| | CCCCTCGCA | TTCGCGGAGC | CGGGACACGA | GAATTGGAGC | TGGCCATTAG | AATCACTCTT | 180 |
| | TACGCAGTGA | TCTTCCTGAT | GAGCGTTGGA | GGAAATATGC | TCATCATCGT | GGTCCTGGGA | 240 |
| | CTGAGCCGCC | GCCTGAGGAC | TGTCACCAAT | GCCTTCCTCC | TCTCACTGGC | AGTCAGCGAC | 300 |
| | CTCCTGCTGG | CTGTGGCTTG | CATGCCCTTC | ACCCTCCTGC | CCAATCTCAT | GGGCACATTC | 360 |
| | ATCTTTGGCA | CCGTCATCTG | CAAGGCGGTT | TCCTACCTCA | TGGGGGTGTC | TGTGAGTGTG | 420 |
| | TCCACGCTAA | GCCTCGTGGC | CATCGCACTG | GAGCGATATA | GCGCCATCTG | CCGACCACTG | 480 |
| | CAGGCACGAG | TGTGGCAGAC | GCGCTCCCAC | GCGGCTCGCG | TGATTGTAGC | CACGTGGCTG | 540 |
| | CTGTCCGGAC | TACTCATGGT | GCCCTACCCC | GTGTACACTG | TCGTGCAACC | AGTGGGGCCT | 600 |
| | CGTGTGCTGC | AGTGCGTGCA | TCGCTGGCCC | AGTGCGCGGG | TCCGCCAGAC | CTGGTCCGTA | 660 |
| | CTGCTGCTTC | TGCTCTTGTT | CTTCATCCCA | GGTGTGGTTA | TGGCCGTGGC | CTACGGGCTT | 720 |
|) | ATCTCTCGCG | AGCTCTACTT | AGGGCTTCGC | TTTGACGGCG | ACAGTGACAG | CGACAGCCAA | 780 |
| | AGCAGGGTCC | GAAACCAAGG | CGGGCTGCCA | GGGGCTGTTC | ACCAGAACGG | GCGTTGCCGG | 840 |

15

| CCTGAGACTG | GCGCGGTTGG | CAAAGACAGC | GATGGCTGCT | ACGTGCAACT | TCCACGTTCC | 900 |
|------------|------------|------------|------------|------------|------------|------|
| CGGCCTGCCC | TGGAGCTGAC | GGCGCTGACG | GCTCCTGGGC | CGGGATCCGG | CTCCCGGCCC | 960 |
| ACCCAGGCCA | AGCTGCTGGC | TAAGAAGCGC | GTGGTGCGAA | TGTTGCTGGT | GATCGTTGTG | 1020 |
| CTTTTTTTC | TGTGTTGGTT | GCCAGTTTAT | AGTGCCAACA | CGTGGCGCGC | CTTTGATGGC | 1080 |
| CCGGGTGCAC | ACCGAGCACT | CTCGGGTGCT | CCTATCTCCT | TCATTCACTT | GCTGAGCTAC | 1140 |
| GCCTCGGCCT | GTGTCAACCC | CCTGGTCTAC | TGCTTCATGC | ACCGTCGCTT | TCGCCAGGCC | 1200 |
| TGCCTGGAAA | CTTGCGCTCG | CTGCTGCCCC | CGGCCTCCAC | GAGCTCGCCC | CAGGGCTCTT | 1260 |
| CCCGATGAGG | ACCCTCCCAC | TCCCTCCATT | GCTTCGCTGT | CCAGGCTTAG | CTACACCACC | 1320 |
| ATCAGCACAC | TGGGCCCTGG | CTGA | | | | 1344 |

- 10 (79) INFORMATION FOR SEQ ID NO:78:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 447 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

Met Glu Leu Leu Lys Leu Asn Arg Ser Val Gln Gly Thr Gly Pro Gly
1 5 10 15

20 Pro Gly Ala Ser Leu Cys Arg Pro Gly Ala Pro Leu Leu Asn Ser Ser 20 25 30

Ser Val Gly Asn Leu Ser Cys Glu Pro Pro Arg Ile Arg Gly Ala Gly 35 40 45

Thr Arg Glu Leu Glu Leu Ala Ile Arg Ile Thr Leu Tyr Ala Val Ile
55 50 55 60

Phe Leu Met Ser Val Gly Gly Asn Met Leu Ile Ile Val Val Leu Gly 65 70 75 80

Leu Ser Arg Arg Leu Arg Thr Val Thr Asn Ala Phe Leu Leu Ser Leu 85 90 95

Ala Val Ser Asp Leu Leu Leu Ala Val Ala Cys Met Pro Phe Thr Leu 100 105 110

Leu Pro Asn Leu Met Gly Thr Phe Ile Phe Gly Thr Val Ile Cys Lys
115 120 125

| | Ala | Val 130 | Ser | Tyr | Leu | Met | Gly 135 | Val | Ser | Val | Ser | Val 140 | Ser | Thr | Leu : | Ser |
|----|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | Leu 145 | Val | Ala | Ile | Ala | Leu 150 | Glu | Arg | Tyr | Ser | Ala 155 | Ile | Cys | Arg | | Leu 160 |
| 5 | Gln | Ala | Arg | Val | Trp 165 | Gln | Thr | Arg | Ser | His 170 | Ala | Ala | Arg | Val | Ile 175 | Val |
| | Ala | Thr | Trp | Leu 180 | Leu | Ser | Gly | Leu | Leu 185 | Met | Val | Pro | Tyr | Pro 190 | Val | Tyr |
| 10 | Thr | Val | Val 195 | Gln | Pro | Val | Gly | Pro 200 | Arg | Val | Leu | Gln | Cys 205 | Val | His | Arg |
| | _ | 210 | | | | | 215 | | | | | 220 | | | Leu | |
| | 225 | | | | | 230 | | | | | 235 | | | | Gly | 240 |
| 15 | | | | | 245 | | | | | 250 | | | | | Ser 255 | |
| | | | | 260 | | | | | 265 | | | | | 270 | Gly | |
| 20 | | | 275 | | | | | 280 | | | | | 285 | | Gly | |
| | | 290 | | | | | 295 | | | | | 300 | | | Ala | |
| | 305 | | | | | 310 | | | | | 315 | | | | | Pro 320 |
| 25 | | | | | 325 | i | | | | 330 |) | | | | 335 | |
| | | | | 340 | | | | | 345 | 5 | | | | 350 | | Ala |
| 30 | | | 355 | ; | | | | 360 |) | | | | 365 | 5 | | Ser |
| | | 370 |) | | | | 375 | 5 | | | | 380 | 0 | | | Cys |
| | 385 | 5 | | | | 390 | ס | | | | 39! | 5 | | | | Ala 400 |
| 35 | | | | | 40 | 5 | | | | 41 | 0 | | | | 415 | |
| | Pro | o Ar | g Ala | a Lei | ı Pr | o As | p Gl | u Asj | p Pr | o Pr | o Th | r Pr | o se | I 110 | = Ala | a Ser |

440

420

435

425

Leu Ser Arg Leu Ser Tyr Thr Thr Ile Ser Thr Leu Gly Pro Gly

| | (80) INFORMATION FOR SEQ ID NO:79: | |
|----|--|-----|
| 5 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| 10 | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:79: | |
| | TGCAAGCTTA AAAAGGAAAA AATGAACAGC | 30 |
| | (81) INFORMATION FOR SEQ ID NO:80: | - |
| 15 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| 20 | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:80: | |
| | TAAGGATCCC TTCCCTTCAA AACATCCTTG | 30 |
| | (82) INFORMATION FOR SEQ ID NO:81: | |
| 25 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1014 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:81: | |
| 30 | ATGAACAGCA CATGTATTGA AGAACAGCAT GACCTGGATC ACTATTTGTT TCCCATTGTT | 60 |
| | TACATCTTTG TGATTATAGT CAGCATTCCA GCCAATATTG GATCTCTGTG TGTGTCTTTC | 120 |
| | CTGCAACCCA AGAAGGAAAG TGAACTAGGA ATTTACCTCT TCAGTTTGTC ACTATCAGAT | 180 |
| | TTACTCTATG CATTAACTCT CCCTTTATGG ATTGATTATA CTTGGAATAA AGACAACTGG | 240 |
| | | |

| | ACTTTCTCTC | CTGCCTTGTG | CAAAGGGAGT | GCTTTTCTCA | TGTACATGAA | GTTTTACAGC | 300 |
|----|------------|------------|------------|------------|------------|------------|------|
| | AGCACAGCAT | TCCTCACCTG | CATTGCCGTT | GATCGGTATT | TGGCTGTTGT | CTACCCTTTG | 360 |
| | AAGTTTTTT | TCCTAAGGAC | AAGAAGAATT | GCACTCATGG | TCAGCCTGTC | CATCTGGATA | 420 |
| | TTGGAAACCA | TCTTCAATGC | TGTCATGTTG | TGGGAAGATG | AAACAGTTGT | TGAATATTGC | 480 |
| 5 | GATGCCGAAA | AGTCTAATTT | TACTTTATGC | TATGACAAAT | ACCCTTTAGA | GAAATGGCAA | 540 |
| | ATCAACCTCA | ACTTGTTCAG | GACGTGTACA | GGCTATGCAA | TACCTTTGGT | CACCATCCTG | 600 |
| | ATCTGTAACC | GGAAAGTCTA | CCAAGCTGTG | CGGCACAATA | AAGCCACGGA | AAACAAGGAA | 660 |
| | AAGAAGAGAA | TCATAAAACT | ACTTGTCAGC | ATCACAGTTA | CTTTTGTCTT | ATGCTTTACT | 720 |
| | CCCTTTCATG | TGATGTTGCT | GATTCGCTGC | ATTTTAGAGC | ATGCTGTGAA | CTTCGAAGAC | 780 |
| 10 | CACAGCAATT | CTGGGAAGCG | AACTTACACA | ATGTATAGAA | TCACGGTTGC | ATTAACAAGT | 840 |
| | TTAAATTGTG | TTGCTGATCC | AATTCTGTAC | TGTTTTGTTA | CCGAAACAGG | AAGATATGAT | 900 |
| | ATGTGGAATA | TATTAAAATT | CTGCACTGGG | AGGTGTAATA | CATCACAAAG | ACAAAGAAAA | 960 |
| | CGCATACTTT | CTGTGTCTAC | AAAAGATACT | ATGGAATTAG | AGGTCCTTGA | GTAG | 1014 |

- (83) INFORMATION FOR SEQ ID NO:82:
- 15 (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 337 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- 20 (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Met Asn Ser Thr Cys Ile Glu Glu Gln His Asp Leu Asp His Tyr Leu 1 5 10 15

Phe Pro Ile Val Tyr Ile Phe Val Ile Ile Val Ser Ile Pro Ala Asn 25 20 25 30

Ile Gly Ser Leu Cys Val Ser Phe Leu Gln Pro Lys Lys Glu Ser Glu 35 40 45

Leu Gly Ile Tyr Leu Phe Ser Leu Ser Leu Ser Asp Leu Leu Tyr Ala 50 55 60

Leu Thr Leu Pro Leu Trp Ile Asp Tyr Thr Trp Asn Lys Asp Asn Trp 65 70 75 80

Thr Phe Ser Pro Ala Leu Cys Lys Gly Ser Ala Phe Leu Met Tyr Met

Glu

30

1 1 I

٠...

Į,

The state of

- (84) INFORMATION FOR SEQ ID NO:83:
- (i) SEOUENCE CHARACTERISTICS: 35
 - (A) LENGTH: 40 base pairs
 - (B) TYPE: nucleic acid

20

30

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:
- 5 CAGGAAGAAG AAACGAGCTG TCATTATGAT GGTGACAGTG 40
 - (85) INFORMATION FOR SEQ ID NO:84:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 40 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:
- 15 CACTGTCACC ATCATAATGA CAGCTCGTTT CTTCTTCCTG
 40
 - (86) INFORMATION FOR SEQ ID NO:85:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:
- 25 GGCCACCGGC AGACCAAACG CGTCCTGCTG
 30
 - (87) INFORMATION FOR SEQ ID NO:86:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

AREN-0054

| | (88) INFORMATION FOR SEQ 1D NO:87: | |
|----|--|-----|
| 5 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 37 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| 10 | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:87: | |
| | GGAAAAGAAG AGAATCAAAA AACTACTTGT CAGCATC | 37 |
| | (89) INFORMATION FOR SEQ ID NO:88: | |
| 15 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:88: | |
| 20 | CTCCTTCGGT CCTCCTATCG TTGTCAGAAG T | 31 |
| | (90) INFORMATION FOR SEQ ID NO:89: | |
| 25 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1080 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:89: | |
| | ATGATTCTCA ACTCTTCTAC TGAAGATGGT ATTAAAAGAA TCCAAGATGA TTGTCCCAAA | 60 |
| 30 | GCTGGAAGGC ATAATTACAT ATTTGTCATG ATTCCTACTT TATACAGTAT CATCTTTGTG | 120 |
| | GTGGGAATAT TTGGAAACAG CTTGGTGGTG ATAGTCATTT ACTTTTATAT GAAGCTGAAG | 180 |
| | ACTGTGGCCA GTGTTTTCT TTTGAATTTA GCACTGGCTG ACTTATGCTT TTTACTGACT | 240 |
| | TTGCCACTAT GGGCTGTCTA CACAGCTATG GAATACCGCT GGCCCTTTGG CAATTACCTA | 300 |

- 132 -

CTCCTTCGGT CCTCCTATCG TTGTCAGAAG T

PATENT

| á | 227 | |
|---------|--------------------|---------|
| ą | === | i |
| | | |
| ä | Ï | |
| ÷ | 1 | |
| | T T | |
| 1 | gudi Budi | 1 |
| | 1 | 1111111 |
| 1111111 | Į. | CHILL. |
| 11 | | |
| | ### ### | |
| i | Ţ | |
| i | 21 3110 2110 | |
| ÷ | ÷., | 7000 |
| | 2710 2710 | |
| | | |

| | TGTAAGATTG | CTTCAGCCAG | CGTCAGTTTC | AACCTGTACG | CTAGTGTGTT | TCTACTCACG | 360 |
|----|------------|------------|------------|------------|------------|------------|------|
| | TGTCTCAGCA | TTGATCGATA | CCTGGCTATT | GTTCACCCAA | TGAAGTCCCG | CCTTCGACGC | 420 |
| | ACAATGCTTG | TAGCCAAAGT | CACCTGCATC | ATCATTTGGC | TGCTGGCAGG | CTTGGCCAGT | 480 |
| | TTGCCAGCTA | TAATCCATCG | AAATGTATTT | TTCATTGAGA | ACACCAATAT | TACAGTTTGT | 540 |
| 5 | GCTTTCCATT | ATGAGTCCCA | AAATTCAACC | CTTCCGATAG | GGCTGGGCCT | GACCAAAAAT | 600 |
| | ATACTGGGTT | TCCTGTTTCC | TTTTCTGATC | ATTCTTACAA | GTTATACTCT | TATTTGGAAG | 660 |
| | GCCCTAAAGA | AGGCTTATGA | AATTCAGAAG | AACAAACCAA | GAAATGATGA | TATTAAAAAG | 720 |
| | ATAATTATGG | CAATTGTGCT | TTTCTTTTTC | TTTTCCTGGA | TTCCCCACCA | AATATTCACT | 780 |
| | TTTCTGGATG | TATTGATTCA | ACTAGGCATC | ATACGTGACT | GTAGAATTGC | AGATATTGTG | 840 |
| 10 | GACACGGCCA | TGCCTATCAC | CATTTGTATA | GCTTATTTTA | ACAATTGCCT | GAATCCTCTT | 900 |
| | TTTTATGGCT | TTCTGGGGAA | AAAATTTAAA | AGATATTTC | TCCAGCTTCT | TTATATAAAA | 960 |
| | CCCCCAAAAG | CCAAATCCCA | CTCAAACCTT | TCAACAAAAA | TGAGCACGCT | TTCCTACCGC | 1020 |
| | CCCTCAGATA | ATGTAAGCTC | ATCCACCAAG | AAGCCTGCAC | CATGTTTTGA | GGTTGAGTGA | 1080 |

- (91) INFORMATION FOR SEQ ID NO:90:
- (i) SEQUENCE CHARACTERISTICS: 15
 - (A) LENGTH: 359 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein 20
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

Met Ile Leu Asn Ser Ser Thr Glu Asp Gly Ile Lys Arg Ile Gln Asp

Asp Cys Pro Lys Ala Gly Arg His Asn Tyr Ile Phe Val Met Ile Pro 25 25 20

Thr Leu Tyr Ser Ile Ile Phe Val Val Gly Ile Phe Gly Asn Ser Leu

Val Val Ile Val Ile Tyr Phe Tyr Met Lys Leu Lys Thr Val Ala Ser

Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Leu Cys Phe Leu Leu Thr 30 75 70 65

Leu Pro Leu Trp Ala Val Tyr Thr Ala Met Glu Tyr Arg Trp Pro Phe

| | | | | | | | _ | - | | | | | | | * - | LE ISIN |
|----|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | | | | | 85 | | | | | 90 | | | | | 95 | |
| | Gly | Asn | Tyr | Leu 100 | Cys | Lys | Ile | Ala | Ser 105 | Ala | Ser | Val | Ser | Phe 110 | Asn | Leu |
| 5 | Tyr | Ala | Ser 115 | Val | Phe | Leu | Leu | Thr 120 | Cys | Leu | Ser | Ile | Asp 125 | Arg | Tyr | Leu |
| | Ala | Ile 130 | Val | His | Pro | Met | Lys 135 | Ser | Arg | Leu | Arg | Arg 140 | Thr | Met | Leu | Val |
| | Ala 145 | Lys | Val | Thr | Cys | Ile 150 | Ile | Ile | Trp | Leu | Leu 155 | Ala | Gly | Leu | Ala | Ser 160 |
| 10 | Leu | Pro | Ala | Ile | Ile 165 | His | Arg | Asn | Val | Phe 170 | Phe | Ile | Glu | Asn | Thr 175 | Asn |
| | Ile | Thr | Val | Cys 180 | Ala | Phe | His | Tyr | Glu 185 | Ser | Gln | Asn | Ser | Thr 190 | Leu | Pro |
| 15 | Ile | Gly | Leu 195 | Gly | Leu | Thr | Lys | Asn 200 | Ile | Leu | Gly | Phe | Leu 205 | Phe | Pro | Phe |
| | Leu | Ile 210 | Ile | Leu | Thr | Ser | Tyr 215 | Thr | Leu | Ile | Trp | Lys 220 | Ala | Leu | Lys | Lys |
| | Ala 225 | Tyr | Glu | Ile | Gln | Lys 230 | Asn | Lys | Pro | Arg | Asn 235 | Asp | Asp | Ile | Lys | Lys 240 |
| 20 | Ile | Ile | Met | Ala | Ile 245 | Val | Leu | Phe | Phe | Phe 250 | Phe | Ser | Trp | Ile | Pro 255 | His |
| | Gln | Ile | Phe | Thr 260 | Phe | Leu | Asp | Val | Leu 265 | Ile | Gln | Leu | Gly | Ile 270 | Ile | Arg |
| 25 | Asp | Cys | Arg 275 | Ile | Ala | Asp | Ile | Val 280 | Asp | Thr | Ala | Met | Pro 285 | Ile | Thr | Ile |
| | Cys | Ile 290 | Ala | Tyr | Phe | Asn | Asn 295 | Cys | Leu | Asn | Pro | Leu 300 | Phe | Tyr | Gly | Phe |
| | Leu 305 | Gly | Lys | Lys | Phe | Lys 310 | Arg | Tyr | Phe | Leu | Gln 315 | Leu | Leu | Lys | Tyr | Ile 320 |
| 30 | Pro | Pro | Lys | Ala | Lys 325 | Ser | His | Ser | Asn | Leu 330 | Ser | Thr | Lys | Met | Ser 335 | Thr |
| | Leu | Ser | Tyr | Arg 340 | Pro | Ser | Asp | Asn | Val 345 | Ser | Ser | Ser | Thr | Lys 350 | Lys | Pro |
| 35 | Ala | Pro | Cys 355 | Phe | Glu | Val | Glu | | | | | | | | | |

(92) INFORMATION FOR SEQ ID NO:91:

AREN-0054

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 35 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single 5 (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:91: 35 31 (ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:93: 25 ATGATTCTCA ACTCTTCTAC TGAAGATGGT ATTAAAAGAA TCCAAGATGA TTGTCCCAAA 60 GCTGGAAGGC ATAATTACAT ATTTGTCATG ATTCCTACTT TATACAGTAT CATCTTTGTG GTGGGAATAT TTGGAAACAG CTTGGTGGTG ATAGTCATTT ACTTTTATAT GAAGCTGAAG 180 ACTGTGGCCA GTGTTTTCT TTTGAATTTA GCACTGGCTG ACTTATGCTT TTTACTGACT 240 30 TTGCCACTAT GGGCTGTCTA CACAGCTATG GAATACCGCT GGCCCTTTGG CAATTACCTA 300 TGTAAGATTG CTTCAGCCAG CGTCAGTTTC GCCCTGTACG CTAGTGTGTT TCTACTCACG 360

TGTCTCAGCA TTGATCGATA CCTGGCTATT GTTCACCCAA TGAAGTCCCG CCTTCGACGC

- 135 -

PATENT

| | ACAATGCTTG | TAGCCAAAGT | CACCTGCATC | ATCATTTGGC | TGCTGGCAGG | CTTGGCCAGT | 480 |
|----|------------|------------|------------|------------|------------|------------|------|
| | TTGCCAGCTA | TAATCCATCG | AAATGTATTT | TTCATTGAGA | ACACCAATAT | TACAGTTTGT | 540 |
| | GCTTTCCATT | ATGAGTCCCA | AAATTCAACC | CTTCCGATAG | GGCTGGGCCT | GACCAAAAAT | 600 |
| | ATACTGGGTT | TCCTGTTTCC | TTTTCTGATC | ATTCTTACAA | GTTATACTCT | TATTTGGAAG | 660 |
| 5 | GCCCTAAAGA | AGGCTTATGA | AATTCAGAAG | AACAAACCAA | GAAATGATGA | TATTTTTAAG | 720 |
| | ATAATTATGG | CAATTGTGCT | TTTCTTTTTC | TTTTCCTGGA | TTCCCCACCA | AATATTCACT | 780 |
| | TTTCTGGATG | TATTGATTCA | ACTAGGCATC | ATACGTGACT | GTAGAATTGC | AGATATTGTG | 840 |
| | GACACGGCCA | TGCCTATCAC | CATTTGTATA | GCTTATTTTA | ACAATTGCCT | GAATCCTCTT | 900 |
| | TTTTATGGCT | TTCTGGGGAA | AAAATTTAAA | AGATATTTTC | TCCAGCTTCT | AAAATATATT | 960 |
| 10 | CCCCCAAAAG | CCAAATCCCA | CTCAAACCTT | TCAACAAAAA | TGAGCACGCT | TTCCTACCGC | 1020 |
| | CCCTCAGATA | ATGTAAGCTC | ATCCACCAAG | AAGCCTGCAC | CATGTTTTGA | GGTTGAGTGA | 1080 |

(95) INFORMATION FOR SEQ ID NO:94:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 359 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:
- Met Ile Leu Asn Ser Ser Thr Glu Asp Gly Ile Lys Arg Ile Gln Asp 1 5 10 15
 - Asp Cys Pro Lys Ala Gly Arg His Asn Tyr Ile Phe Val Met Ile Pro 20 25 30
- Thr Leu Tyr Ser Ile Ile Phe Val Val Gly Ile Phe Gly Asn Ser Leu 35 40 45
 - Val Val Ile Val Ile Tyr Phe Tyr Met Lys Leu Lys Thr Val Ala Ser 50 55 60
 - Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Leu Cys Phe Leu Leu Thr 65 70 75 80
- Leu Pro Leu Trp Ala Val Tyr Thr Ala Met Glu Tyr Arg Trp Pro Phe 85 90 95
 - Gly Asn Tyr Leu Cys Lys Ile Ala Ser Ala Ser Val Ser Phe Ala Leu

100 105 110 Tyr Ala Ser Val Phe Leu Leu Thr Cys Leu Ser Ile Asp Arg Tyr Leu 120 Ala Ile Val His Pro Met Lys Ser Arg Leu Arg Arg Thr Met Leu Val 5 135 140 Ala Lys Val Thr Cys Ile Ile Ile Trp Leu Leu Ala Gly Leu Ala Ser 150 155 Leu Pro Ala Ile Ile His Arg Asn Val Phe Phe Ile Glu Asn Thr Asn 165 170 Ile Thr Val Cys Ala Phe His Tyr Glu Ser Gln Asn Ser Thr Leu Pro 10 185 Ile Gly Leu Gly Leu Thr Lys Asn Ile Leu Gly Phe Leu Phe Pro Phe Leu Ile Ile Leu Thr Ser Tyr Thr Leu Ile Trp Lys Ala Leu Lys Lys 15 215 Ala Tyr Glu Ile Gln Lys Asn Lys Pro Arg Asn Asp Asp Ile Phe Lys Ile Ile Met Ala Ile Val Leu Phe Phe Phe Phe Ser Trp Ile Pro His 250 Gln Ile Phe Thr Phe Leu Asp Val Leu Ile Gln Leu Gly Ile Ile Arg 20 260 265 Asp Cys Arg Ile Ala Asp Ile Val Asp Thr Ala Met Pro Ile Thr Ile 280 Cys Ile Ala Tyr Phe Asn Asn Cys Leu Asn Pro Leu Phe Tyr Gly Phe 25 295 Leu Gly Lys Lys Phe Lys Arg Tyr Phe Leu Gln Leu Leu Lys Tyr Ile 305 310 Pro Pro Lys Ala Lys Ser His Ser Asn Leu Ser Thr Lys Met Ser Thr 30 Leu Ser Tyr Arg Pro Ser Asp Asn Val Ser Ser Ser Thr Lys Lys Pro 345 350 340 Ala Pro Cys Phe Glu Val Glu 355

(97) INFORMATION FOR SEQ ID NO:95:

35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs

(B) TYPE: nucleic acid

AREN-0054

| | (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
|----|--|----|
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: NO | |
| 5 | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:95: | |
| | CCCAAGCTTC CCCAGGTGTA TTTGAT | 26 |
| | (97) INFORMATION FOR SEQ ID NO:96: | |
| 10 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 29 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: YES | : |
| | | |
| 15 | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96: | |
| | CCTGCAGGCG AAACTGACTC TGGCTGAAG | 29 |
| | (98) INFORMATION FOR SEQ ID NO:97: | |
| 20 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 42 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: NO | |
| 25 | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:97: | |
| | CTGTACGCTA GTGTGTTCT ACTCACGTGT CTCAGCATTG AT | 42 |
| | (99) INFORMATION FOR SEQ ID NO:98: | |
| 30 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 26 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |

(ii) MOLECULE TYPE: DNA (genomic)

The same of the sa

(iv) ANTI-SENSE: YES

| 1231 | SECTIENCE | DESCRIPTION: | SEO | ID | NO:98: |
|------|-----------|--------------|-----|----|--------|

GTTGGATCCA CATAATGCAT TTTCTC

26

(100) INFORMATION FOR SEQ ID NO:99:

- 5 (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1080 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- 10 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

ATGATTCTCA ACTCTTCTAC TGAAGATGGT ATTAAAAGAA TCCAAGATGA TTGTCCCAAA GCTGGAAGGC ATAATTACAT ATTTGTCATG ATTCCTACTT TATACAGTAT CATCTTTGTG 120 GTGGGAATAT TTGGAAACAG CTTGGTGGTG ATAGTCATTT ACTTTTATAT GAAGCTGAAG 180 15 ACTGTGGCCA GTGTTTTCT TTTGAATTTA GCACTGGCTG ACTTATGCTT TTTACTGACT 240 TTGCCACTAT GGGCTGTCTA CACAGCTATG GAATACCGCT GGCCCTTTGG CAATTACCTA 300 TGTAAGATTG CTTCAGCCAG CGTCAGTTTC AACCTGTACG CTAGTGTGTT TCTACTCACG 360 TGTCTCAGCA TTGATCGATA CCTGGCTATT GTTCACCCAA TGAAGTCCCG CCTTCGACGC 420 ACAATGCTTG TAGCCAAAGT CACCTGCATC ATCATTTGGC TGCTGGCAGG CTTGGCCAGT 480 20 TTGCCAGCTA TAATCCATCG AAATGTATTT TTCATTGAGA ACACCAATAT TACAGTTTGT 540 GCTTTCCATT ATGAGTCCCA AAATTCAACC CTTCCGATAG GGCTGGGCCT GACCAAAAAT 600 ATACTGGGTT TCCTGTTTCC TTTTCTGATC ATTCTTACAA GTTATTTTGG AATTCGAAAA 660 CACTTACTGA AGACGAATAG CTATGGGAAG AACAGGATAA CCCGTGACCA AGTTAAGAAG 720 ATAATTATGG CAATTGTGCT TTTCTTTTC TTTTCCTGGA TTCCCCACCA AATATTCACT 780 25 TTTCTGGATG TATTGATTCA ACTAGGCATC ATACGTGACT GTAGAATTGC AGATATTGTG 840 GACACGGCCA TGCCTATCAC CATTTGTATA GCTTATTTTA ACAATTGCCT GAATCCTCTT 900 TTTTATGGCT TTCTGGGGAA AAAATTTAAA AGATATTTC TCCAGCTTCT AAAATATATT 960 CCCCCAAAAG CCAAATCCCA CTCAAACCTT TCAACAAAAA TGAGCACGCT TTCCTACCGC 1020 CCCTCAGATA ATGTAAGCTC ATCCACCAAG AAGCCTGCAC CATGTTTTGA GGTTGAGTGA 1080

(101) INFORMATION FOR SEQ ID NO:100:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 359 amino acids
 - (B) TYPE: amino acid
- (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

| | Met | Ile | Leu | Asn | Ser | Ser | Thr | Glu | Asp | Gly | Ile | Lys | Arg | Ile | Gln | Asp |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 10 | 1 | | | | 5 | | | | | 10 | | | | | 15 | |

Asp Cys Pro Lys Ala Gly Arg His Asn Tyr Ile Phe Val Met Ile Pro 20 25 30

Thr Leu Tyr Ser Ile Ile Phe Val Val Gly Ile Phe Gly Asn Ser Leu 35 40 45

Val Val Ile Val Ile Tyr Phe Tyr Met Lys Leu Lys Thr Val Ala Ser
50 55 60

Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Leu Cys Phe Leu Leu Thr 65 70 75 80

Leu Pro Leu Trp Ala Val Tyr Thr Ala Met Glu Tyr Arg Trp Pro Phe 85 90 95

Gly Asn Tyr Leu Cys Lys Ile Ala Ser Ala Ser Val Ser Phe Asn Leu 100 105 110

Tyr Ala Ser Val Phe Leu Leu Thr Cys Leu Ser Ile Asp Arg Tyr Leu 115 120 125

25 Ala Ile Val His Pro Met Lys Ser Arg Leu Arg Arg Thr Met Leu Val 130 135 140

Leu Pro Ala Ile Ile His Arg Asn Val Phe Phe Ile Glu Asn Thr Asn 30 165 170 175

Ile Thr Val Cys Ala Phe His Tyr Glu Ser Gln Asn Ser Thr Leu Pro 180 185 190

Ile Gly Leu Gly Leu Thr Lys Asn Ile Leu Gly Phe Leu Phe Pro Phe 195 200 205

Leu Ile Ile Leu Thr Ser Tyr Phe Gly Ile Arg Lys His Leu Leu Lys 210 215 220

| ARE | EN-0054 | | - 141 - | | | | | | | | | | PATENT | | | | |
|-----|------------|------------|---|------------------------|------------------------|------------------------|------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|----|
| | Thr 225 | Asn | Ser | Tyr | Gly | Lys 230 | Asn | Arg | Ile | Thr | Arg 235 | Asp | Gln | Val | Lys | Lys 240 | |
| | Ile | Ile | Met | Ala | Ile 245 | Val | Leu | Phe | Phe | Phe 250 | Phe | Ser | Trp | Ile | Pro 255 | His | |
| 5 | Gln | Ile | Phe | Thr 260 | Phe | Leu | Asp | Val | Leu 265 | Ile | Gln | Leu | Gly | Ile 270 | Ile | Arg | |
| | Asp | Cys | Arg 275 | Ile | Ala | Asp | Ile | Val 280 | Asp | Thr | Ala | Met | Pro 285 | Ile | Thr | Ile | |
| 10 | Cys | Ile 290 | Ala | Tyr | Phe | Asn | Asn 295 | Cys | Leu | Asn | Pro | Leu 300 | Phe | Tyr | Gly | Phe | |
| | Leu 305 | Gly | Lys | Lys | Phe | Lys 310 | Arg | Tyr | Phe | Leu | Gln 315 | Leu | Leu | Lys | Tyr | Ile 320 | |
| | Pro | Pro | Lys | Ala | Lys 325 | Ser | His | Ser | Asn | Leu 330 | Ser | Thr | Lys | Met | Ser 335 | Thr | |
| 15 | Leu | Ser | Tyr | Arg 340 | Pro | Ser | Asp | Asn | Val 345 | Ser | Ser | Ser | Thr | Lys 350 | Lys | Pro | |
| | Ala | Pro | Cys 355 | Phe | Glu | Val | Glu | | | | | | | | | | |
| | (102) IN | FORM | ATIO | v FOI | R SE | Q ID | NO: | 101: | | | | | | | | | |
| 20 | (1) | (B | UENCI) LEI) TYI) STI) TOI | NGTH PE: 1 RANDI | : 37 nucle EDNE: | base eic a SS: s | e pa: acid sing: | irs | | | | | | | | | |
| 25 | (ii) | MOL | ECULI | E TY | PE: 1 | ANC | (gend | omic) |) | | | | | | | | |
| | (iv) | ANT | I-SE | NSE: | YES | | | | | | | | | | | | |
| | (xi) | SEQ | UENC | E DES | SCRI | PTIO | N: SI | EQ II | ои с | :101 | : | | | | | | |
| | TCCGAATI | CC A | AAAT | AACT | r GT | AAGA | ATGA | TCA | GAAA | | | | | | | | 37 |
| | (103) IN | FORM | ATIO | v FOI | R SE | Q ID | NO: | 102: | | | | | | | | | |
| 30 | (i) | (B | UENCI) LEI) TYI) STI) TOI | NGTH PE: 1 RANDI | : 33 nucle EDNE | base eic a | e pa: acid sing! | irs | | | | | | | | | |
| 35 | (ii) | MOL | ECULI | E TYI | PE: 1 | ONA | (gend | omic |) | | | | | | | | |

(iv) ANTI-SENSE: NO

| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:102: | |
|----|--|----|
| | AGATCTTAAG AAGATAATTA TGGCAATTGT GCT | 33 |
| | (104) INFORMATION FOR SEQ ID NO:103: | |
| 5 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 62 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| 10 | (iv) ANTI-SENSE: NO | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:103: | |
| | AATTCGAAAA CACTTACTGA AGACGAATAG CTATGGGAAG AACAGGATAA CCCGTGACCA | 60 |
| | AG | 62 |
| | (105) INFORMATION FOR SEQ ID NO:104: | |
| 15 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 62 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| 20 | (ii) MOLECULE TYPE: DNA (genomic) | |
| • | (iv) ANTI-SENSE: YES | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:104: | |
| | TTAACTTGGT CACGGGTTAT CCTGTTCTTC CCATAGCTAT TCGTCTTCAG TAAGTGTTTT | 60 |
| | CG | 62 |
| 25 | (106) INFORMATION FOR SEQ ID NO:105: | |
| 30 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1083 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (ad) GEOMETER DESCRIPTION, SEC ID NO.105. | |

| | ATGATTCTCA | ACTCTTCTAC | TGAAGATGGT | ATTAAAAGAA | TCCAAGATGA | TTGTCCCAAA | 60 |
|----|------------|------------|------------|------------|------------|------------|------|
| | GCTGGAAGGC | ATAATTACAT | ATTTGTCATG | ATTCCTACTT | TATACAGTAT | CATCTTTGTG | 120 |
| | GTGGGAATAT | TTGGAAACAG | CTTGGTGGTG | ATAGTCATTT | ACTTTTATAT | GAAGCTGAAG | 180 |
| | ACTGTGGCCA | GTGTTTTTCT | TTTGAATTTA | GCACTGGCTG | ACTTATGCTT | TTTACTGACT | 240 |
| 5 | TTGCCACTAT | GGGCTGTCTA | CACAGCTATG | GAATACCGCT | GGCCCTTTGG | CAATTACCTA | 300 |
| | TGTAAGATTG | CTTCAGCCAG | CGTCAGTTTC | AACCTGTACG | CTAGTGTGTT | TCTACTCACG | 360 |
| | TGTCTCAGCA | TTGATCGATA | CCTGGCTATT | GTTCACCCAA | TGAAGTCCCG | CCTTCGACGC | 420 |
| | ACAATGCTTG | TAGCCAAAGT | CACCTGCATC | ATCATTTGGC | TGCTGGCAGG | CTTGGCCAGT | 480 |
| | TTGCCAGCTA | TAATCCATCG | AAATGTATTT | TTCATTGAGA | ACACCAATAT | TACAGTTTGT | 540 |
| 10 | GCTTTCCATT | ATGAGTCCCA | AAATTCAACC | CTTCCGATAG | GGCTGGGCCT | GACCAAAAAT | 600 |
| | ATACTGGGTT | TCCTGTTTCC | TTTTCTGATC | ATTCTTACAA | GTTATACTCT | TATTTGGAAG | 660 |
| | GCCCTAAAGA | AGGCTTATGA | AATTCAGAAG | AACAAACCAA | GAAATGATGA | TATTTTTAAG | 720 |
| | ATAATTATGG | CAGCAATTGT | GCTTTTCTTT | TTCTTTTCCT | GGATTCCCCA | CCAAATATTC | 780 |
| | ACTTTTCTGG | ATGTATTGAT | TCAACTAGGC | ATCATACGTG | ACTGTAGAAT | TGCAGATATT | 840 |
| 15 | GTGGACACGG | CCATGCCTAT | CACCATTTGT | ATAGCTTATT | TTAACAATTG | CCTGAATCCT | 900 |
| | CTTTTTTATG | GCTTTCTGGG | GAAAAAATTT | AAAAGATATT | TTCTCCAGCT | TCTAAAATAT | 960 |
| | ATTCCCCCAA | AAGCCAAATC | CCACTCAAAC | CTTTCAACAA | AAATGAGCAC | GCTTTCCTAC | 1020 |
| | CGCCCTCAG | ATAATGTAAG | CTCATCCACC | AAGAAGCCTG | CACCATGTTT | TGAGGTTGAG | 1080 |
| | TGA | | | | | | 1083 |

- 20 (107) INFORMATION FOR SEQ ID NO:106:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 360 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
- 25 (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

Met Ile Leu Asn Ser Ser Thr Glu Asp Gly Ile Lys Arg Ile Gln Asp 1 5 10 15

30 Asp Cys Pro Lys Ala Gly Arg His Asn Tyr Ile Phe Val Met Ile Pro

| | | | | 20 | | | | | 25 | | | | | 30 | | |
|----|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | Thr | Leu | Tyr 35 | Ser | Ile | Ile | Phe | Val 40 | Val | Gly | Ile | Phe | Gly 45 | Asn | Ser | Leu |
| 5 | Val | Val 50 | Ile | Val | Ile | Tyr | Phe 55 | Tyr | Met | Lys | Leu | Lys 60 | Thr | Val | Ala | Ser |
| | Val 65 | Phe | Leu | Leu | Asn | Leu 70 | Ala | Leu | Ala | Asp | Leu 75 | Cys | Phe | Leu | Leu | Thr 80 |
| | Leu | Pro | Leu | Trp | Ala 85 | Val | Tyr | Thr | Ala | Met 90 | Glu | Tyr | Arg | Trp | Pro 95 | Phe |
| 10 | Gly | Asn | Tyr | Leu 100 | Cys | Lys | Ile | Ala | Ser 105 | Ala | Ser | Val | Ser | Phe 110 | Asn | Leu |
| | Tyr | Ala | Ser 115 | Val | Phe | Leu | Leu | Thr 120 | Cys | Leu | Ser | Ile | Asp 125 | Arg | Tyr | Leu |
| 15 | Ala | Ile 130 | Val | His | Pro | Met | Lys 135 | Ser | Arg | Leu | Arg | Arg 140 | Thr | Met | Leu | Val |
| | Ala 145 | Lys | Val | Thr | Cys | Ile 150 | Ile | Ile | Trp | Leu | Leu 155 | Ala | Gly | Leu | Ala | Ser 160 |
| | Leu | Pro | Ala | Ile | Ile 165 | His | Arg | Asn | Val | Phe 170 | Phe | Ile | Glu | Asn | Thr 175 | Asn |
| 20 | Ile | Thr | Val | Cys 180 | Ala | Phe | His | Tyr | Glu 185 | Ser | Gln | Asn | Ser | Thr 190 | Leu | Pro |
| | Ile | Gly | Leu 195 | Gly | Leu | Thr | Lys | Asn 200 | Ile | Leu | Gly | Phe | Leu 205 | Phe | Pro | Phe |
| 25 | Leu | Ile 210 | Ile | Leu | Thr | Ser | Tyr 215 | Thr | Leu | Ile | Trp | Lys 220 | Ala | Leu | Lys | Lys |
| | Ala 225 | Tyr | Glu | Ile | Gln | Lys 230 | Asn | Lys | Pro | Arg | Asn 235 | Asp | Asp | Ile | Phe | Lys 240 |
| | Ile | Ile | Met | Ala | Ala 245 | Ile | Val | Leu | Phe | Phe 250 | Phe | Phe | Ser | Trp | Ile 255 | Pro |
| 30 | His | Gln | Ile | Phe 260 | Thr | Phe | Leu | Asp | Val 265 | Leu | Ile | Gln | Leu | Gly 270 | Ile | Ile |
| | Arg | Asp | Cys 275 | Arg | Ile | Ala | Asp | Ile 280 | Val | Asp | Thr | Ala | Met 285 | Pro | Ile | Thr |
| 35 | Ile | Cys 290 | Ile | Ala | Tyr | Phe | Asn 295 | Asn | Cys | Leu | Asn | Pro 300 | Leu | Phe | Tyr | Gly |
| | Phe 305 | Leu | Gly | Lys | Lys | Phe 310 | Lys | Arg | Tyr | Phe | Leu 315 | Gln | Leu | Leu | Lys | Tyr 320 |

Ile Pro Pro Lys Ala Lys Ser His Ser Asn Leu Ser Thr Lys Met Ser 325 330 Thr Leu Ser Tyr Arg Pro Ser Asp Asn Val Ser Ser Ser Thr Lys Lys 345 Pro Ala Pro Cys Phe Glu Val Glu 355 (108) INFORMATION FOR SEQ ID NO:107: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 26 base pairs (B) TYPE: nucleic acid 10 (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (iv) ANTI-SENSE: NO : (xi) SEQUENCE DESCRIPTION: SEQ ID NO:107: 15 CCCAAGCTTC CCCAGGTGTA TTTGAT 26 (109) INFORMATION FOR SEQ ID NO:108: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 38 base pairs (B) TYPE: nucleic acid 20 (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (iv) ANTI-SENSE: YES (xi) SEQUENCE DESCRIPTION: SEQ ID NO:108: 25 38 AAGCACAATT GCTGCATAAT TATCTTAAAA ATATCATC (110) INFORMATION FOR SEQ ID NO:109: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid 30 (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

AREN-0054 - 146 -**PATENT** (xi) SEQUENCE DESCRIPTION: SEQ ID NO:109: AAGATAATTA TGGCAGCAAT TGTGCTTTTC TTTTTCTTT 39 (111) INFORMATION FOR SEQ ID NO:110: (i) SEQUENCE CHARACTERISTICS: 5 (A) LENGTH: 26 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) 10 (iv) ANTI-SENSE: YES (xi) SEQUENCE DESCRIPTION: SEQ ID NO:110: 26 GTTGGATCCA CATAATGCAT TTTCTC (112) INFORMATION FOR SEQ ID NO:111: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1344 base pairs 15 (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:111: 20 ATGGAGCTGC TAAAGCTGAA CCGGAGCGTG CAGGGAACCG GACCCGGGCC GGGGGCTTCC 60 CTGTGCCGCC CGGGGGCGCC TCTCCTCAAC AGCAGCAGTG TGGGCAACCT CAGCTGCGAG 120 CCCCCTCGCA TTCGCGGAGC CGGGACACGA GAATTGGAGC TGGCCATTAG AATCACTCTT 180 TACGCAGTGA TCTTCCTGAT GAGCGTTGGA GGAAATATGC TCATCATCGT GGTCCTGGGA 240

25 CTGAGCCGCC GCCTGAGGAC TGTCACCAAT GCCTTCCTCC TCTCACTGGC AGTCAGCGAC

CTCCTGCTGG CTGTGGCTTG CATGCCCTTC ACCCTCCTGC CCAATCTCAT GGGCACATTC

ATCTTTGGCA CCGTCATCTG CAAGGCGGTT TCCTACCTCA TGGGGGTGTC TGTGAGTGTG

TCCACGCTAA GCCTCGTGGC CATCGCACTG GAGCGATATA GCGCCATCTG CCGACCACTG

CAGGCACGAG TGTGGCAGAC GCGCTCCCAC GCGGCTCGCG TGATTGTAGC CACGTGGCTG

CGTGTGCTGC AGTGCGTGCA TCGCTGGCCC AGTGCGCGGG TCCGCCAGAC CTGGTCCGTA

30 CTGTCCGGAC TACTCATGGT GCCCTACCCC GTGTACACTG TCGTGCAACC AGTGGGGCCT

300

420

480

540

600

660

| AREN-0054 | | | | - 1 | PATENT | | | |
|-----------|----|------------|------------|------------|------------|------------|------------|------|
| | | CTGCTGCTTC | TGCTCTTGTT | CTTCATCCCA | GGTGTGGTTA | TGGCCGTGGC | CTACGGGCTT | 720 |
| | | ATCTCTCGCG | AGCTCTACTT | AGGGCTTCGC | TTTGACGGCG | ACAGTGACAG | CGACAGCCAA | 780 |
| | | AGCAGGGTCC | GAAACCAAGG | CGGGCTGCCA | GGGGCTGTTC | ACCAGAACGG | GCGTTGCCGG | 840 |
| | | CCTGAGACTG | GCGCGGTTGG | CAAAGACAGC | GATGGCTGCT | ACGTGCAACT | TCCACGTTCC | 900 |
| | 5 | CGGCCTGCCC | TGGAGCTGAC | GGCGCTGACG | GCTCCTGGGC | CGGGATCCGG | CTCCCGGCCC | 960 |
| | | ACCCAGGCCA | AGCTGCTGGC | TAAGAAGCGC | GTGAAACGAA | TGTTGCTGGT | GATCGTTGTG | 1020 |
| | | CTTTTTTTC | TGTGTTGGTT | GCCAGTTTAT | AGTGCCAACA | CGTGGCGCGC | CTTTGATGGC | 1080 |
| | | CCGGGTGCAC | ACCGAGCACT | CTCGGGTGCT | CCTATCTCCT | TCATTCACTT | GCTGAGCTAC | 1140 |
| | | GCCTCGGCCT | GTGTCAACCC | CCTGGTCTAC | TGCTTCATGC | ACCGTCGCTT | TCGCCAGGCC | 1200 |
| | 10 | TGCCTGGAAA | CTTGCGCTCG | CTGCTGCCCC | CGGCCTCCAC | GAGCTCGCCC | CAGGGCTCTT | 1260 |
| | | CCCGATGAGG | ACCCTCCCAC | TCCCTCCATT | GCTTCGCTGT | CCAGGCTTAG | CTACACCACC | 1320 |

(113) INFORMATION FOR SEQ ID NO:112:

ATCAGCACAC TGGGCCCTGG CTGA

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 447 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

Met Glu Leu Leu Lys Leu Asn Arg Ser Val Gln Gly Thr Gly Pro Gly
1 5 10 15

1344

Pro Gly Ala Ser Leu Cys Arg Pro Gly Ala Pro Leu Leu Asn Ser Ser 20 25 30

25 Ser Val Gly Asn Leu Ser Cys Glu Pro Pro Arg Ile Arg Gly Ala Gly
35 40 45

Thr Arg Glu Leu Glu Leu Ala Ile Arg Ile Thr Leu Tyr Ala Val Ile 50 55 60

Phe Leu Met Ser Val Gly Gly Asn Met Leu Ile Ile Val Val Leu Gly 30 65 70 75 80

Leu Ser Arg Arg Leu Arg Thr Val Thr Asn Ala Phe Leu Leu Ser Leu 85 90 95

| | Ala | Val | Ser | Asp 100 | Leu | Leu | Leu | Ala | Val 105 | Ala | Cys | Met | Pro | Phe 110 | Thr | Leu |
|----|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | Leu | Pro | Asn 115 | Leu | Met | Gly | Thr | Phe 120 | Ile | Phe | Gly | Thr | Val 125 | Ile | Cys | Lys |
| 5 | Ala | Val 130 | Ser | Tyr | Leu | Met | Gly 135 | Val | Ser | Val | Ser | Val 140 | Ser | Thr | Leu | Ser |
| | Leu 145 | Val | Ala | Ile | Ala | Leu 150 | Glu | Arg | Tyr | Ser | Ala 155 | Ile | Cys | Arg | Pro | Leu 160 |
| 10 | Gln | Ala | Arg | Val | Trp 165 | Gln | Thr | Arg | Ser | His 170 | Ala | Ala | Arg | Val | Ile 175 | Val |
| | Ala | Thr | Trp | Leu 180 | Leu | Ser | Gly | Leu | Leu 185 | Met | Val | Pro | Tyr | Pro 190 | Val | Tyr |
| | Thr | Val | Val 195 | Gln | Pro | Val | Gly | Pro 200 | Arg | Val | Leu | Gln | Cys 205 | Val | His | Arg |
| 15 | Trp | Pro 210 | Ser | Ala | Arg | Val | Arg 215 | Gln | Thr | Trp | Ser | Val 220 | Leu | Leu | Leu | Leu |
| | Leu 225 | Leu | Phe | Phe | Ile | Pro 230 | Gly | Val | Val | Met | Ala 235 | Val | Ala | Tyr | Gly | Leu 240 |
| 20 | Ile | Ser | Arg | Glu | Leu 245 | Tyr | Leu | Gly | Leu | Arg 250 | Phe | Asp | Gly | Asp | Ser 255 | Asp |
| | Ser | Asp | Ser | Gln 260 | Ser | Arg | Val | Arg | Asn 265 | Gln | Gly | Gly | Leu | Pro 270 | Gly | Ala |
| | Val | His | Gln 275 | Asn | Gly | Arg | Cys | Arg 280 | Pro | Glu | Thr | Gly | Ala 285 | Val | Gly | Lys |
| 25 | Asp | Ser 290 | Asp | Gly | Cys | Tyr | Val 295 | Gln | Leu | Pro | Arg | Ser 300 | Arg | Pro | Ala | Leu |
| | Glu 305 | Leu | Thr | Ala | Leu | Thr 310 | Ala | Pro | Gly | Pro | Gly 315 | Ser | Gly | Ser | Arg | Pro 320 |
| 30 | Thr | Gln | Ala | Lys | Leu 325 | | Ala | Lys | Lys | Arg 330 | | Lys | Arg | Met | Leu 335 | Leu |
| | Val | Ile | Val | Val 340 | | Phe | Phe | Leu | Cys 345 | | Leu | Pro | Val | Tyr 350 | Ser | Ala |
| | Asn | Thr | Trp 355 | | Ala | Phe | Asp | Gly 360 | | Gly | Ala | His | Arg 365 | | Leu | Ser |
| 35 | Val | Ala 370 | | Ile | Ser | Phe | 11e 375 | | Leu | Leu | . Ser | Tyr 380 | | Ser | Ala | Cys |
| | Val | Asn | Pro | Leu | Val | Tyr | Cys | Phe | Met | His | Arg | Arg | Phe | arg | Gln | Ala |

ID , ¹ ijħ i. a Fi 11 £ ١.... n

PATENT 400 395 385 390 Cys Leu Glu Thr Cys Ala Arg Cys Cys Pro Arg Pro Pro Arg Ala Arg 410 405 Pro Arg Ala Leu Pro Asp Glu Asp Pro Pro Thr Pro Ser Ile Ala Ser 5 425 Leu Ser Arg Leu Ser Tyr Thr Thr Ile Ser Thr Leu Gly Pro Gly 440 435 (114) INFORMATION FOR SEQ ID NO:113: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 34 base pairs 10 (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:113: 15 34 CAGCAGCATG CGCTTCACGC GCTTCTTAGC CCAG (115) INFORMATION FOR SEQ ID NO:114: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid 20 (C) STRANDEDNESS: single (D) TOPOLOGY: not relevant (ii) MOLECULE TYPE: DNA (genomic) (xi) SEQUENCE DESCRIPTION: SEQ ID NO:114: 35 25 AGAAGCGCGT GAAGCGCATG CTGCTGGTGA TCGTT (116) INFORMATION FOR SEQ ID NO:115: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single 30 (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

AREN-0054

| | ATGGAGAAAA GAATCAAAAG AATGTTCTAT ATA | 33 |
|----|--|----|
| | (117) INFORMATION FOR SEQ ID NO:116: | |
| 5 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: YES | |
| 10 | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:116: | |
| | TATATAGAAC ATTCTTTTGA TTCTTTTCTC CAT | 33 |
| | (118) INFORMATION FOR SEQ ID NO:117: | |
| 15 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | ; |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: NO | |
| 20 | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:117: | |
| | CGCTCTCTGG CCTTGAAGCG CACGCTCAGC | 30 |
| | (119) INFORMATION FOR SEQ ID NO:118: | |
| 25 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: YES | |
| 30 | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:118: | |
| | GCTGAGCGTG CGCTTCAAGG CCAGAGAGCG | 30 |
| | (120) INFORMATION FOR SEQ ID NO:119: | |

- 150 -

PATENT

AREN-0054

| 5 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
|----|--|----|
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: NO | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:119: | |
| | CCCAGGAAAA AGGTGAAAGT CAAAGTTTTC | 30 |
| 10 | (121) INFORMATION FOR SEQ ID NO:120: | |
| 15 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: YES | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:120: | |
| | GAAAACTTTG ACTTTCACCT TTTTCCTGGG | 30 |
| 20 | (122) INFORMATION FOR SEQ ID NO:121: | |
| 25 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: NO | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:121: | |
| | GGGGCGCGGG TGAAACGGCT GGTGAGC | 27 |
| 30 | (123) INFORMATION FOR SEQ ID NO:122: | |
| | (i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 27 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single | |

- 151 -

PATENT

| | (D) TOPOLOGY: linear | |
|----|--|----|
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: YES | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:122: | |
| 5 | GCTCACCAGC CGTTTCACCC GCGCCCC | 27 |
| | (124) INFORMATION FOR SEQ ID NO:123: | |
| 10 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: NO | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:123: | 30 |
| 15 | CCCCTTGAAA AGCCTAAGAA CTTGGTCATC | 30 |
| | (125) INFORMATION FOR SEQ ID NO:124: | |
| 20 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 30 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (iv) ANTI-SENSE: YES | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:124: | |
| 25 | GATGACCAAG TTCTTAGGCT TTTCAAGGGG | 30 |
| | (126) INFORMATION FOR SEQ ID NO:125: | |
| 30 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 32 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |

(ii) MOLECULE TYPE: DNA (genomic)

| - | - | 1 |
|---------------|---|---------|
| | Ē | ••• |
| 1 | 2 | |
| * | 142 | States |
| .000 | Start! | 3,000 |
| 241111 | there's | in the |
| 1111117 | thun | ione. |
| 1111111 | that it | Section |
| - | | |
| 411117 | ======================================= | į |
| ALC: | 77 | i. |
| 7 | ** | į |
| 1 | , II | |
| Title Control | 11 | į |
| 4 | 72 | à |

| (xi) | SEOUENCE | DESCRIPTION: | SEO | ID | NO:125: |
|------|----------|--------------|-----|----|---------|

GATCTCTAGA ATGAACAGCA CATGTATTGA AG

32

- 5 (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 35 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- 10 (ii) MOLECULE TYPE: DNA (genomic)
 - (iv) ANTI-SENSE: YES
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

CTAGGGTACC CGCTCAAGGA CCTCTAATTC CATAG

35

(128) INFORMATION FOR SEQ ID NO:127:

- 15 (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1296 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- 20 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

ATGCAGGCGC TTAACATTAC CCCGGAGCAG TTCTCTCGGC TGCTGCGGGA CCACAACCTG 60 ACGCGGGAGC AGTTCATCGC TCTGTACCGG CTGCGACCGC TCGTCTACAC CCCAGAGCTG 120 CCGGGACGCG CCAAGCTGGC CCTCGTGCTC ACCGGCGTGC TCATCTTCGC CCTGGCGCTC 180 25 TTTGGCAATG CTCTGGTGTT CTACGTGGTG ACCCGCAGCA AGGCCATGCG CACCGTCACC 240 AACATCTTTA TCTGCTCCTT GGCGCTCAGT GACCTGCTCA TCACCTTCTT CTGCATTCCC 300 GTCACCATGC TCCAGAACAT TTCCGACAAC TGGCTGGGGG GTGCTTTCAT TTGCAAGATG 360 GTGCCATTTG TCCAGTCTAC CGCTGTTGTG ACAGAAATGC TCACTATGAC CTGCATTGCT 420 GTGGAAAGGC ACCAGGGACT TGTGCATCCT TTTAAAATGA AGTGGCAATA CACCAACCGA 480

| ARE | N-0054 | | | | - 1 | 54 - | | | | | | | PA | TEN | IT |
|-----|-----------|--|---------------------------|-----------------------------|-------------|-----------|-----------|-----------|--------|--------|-----------|-----------|---------------|-----|------|
| | AGGGCTTTC | A CAATGO | TAGG ' | TGTGGTC | TGG | CTGG | TGGC | AG : | CATO | GTAG | G A | CAC | CCATO | 3 | 540 |
| | TGGCACGTG | C AACAAC | TTGA (| GATCAAA | TAT | GACT | TCCT | AT A | ATGAA | AAGG | A A | CACA | rctgo | 2 | 600 |
| | TGCTTAGAA | G AGTGGA | CCAG | CCCTGTG | CAC | CAGA | AGAT | CT A | ACACC | ACCI | T C | ATCC | rtgtc | 2 | 660 |
| | ATCCTCTTC | C TCCTGC | CTCT ' | TATGGTG | ATG | CTT | TTCT | GT A | ACAGI | 'AAAA' | T TO | GTT | ATGAZ | 4 | 720 |
| 5 | CTTTGGATA | A AGAAAA | GAGT ' | TGGGGAT | GGT | TCAC | TGCT | TC (| GAACT | TATTO | A TO | GAA! | AAGAZ | Ā | 780 |
| | ATGTCCAAA | A TAGCCA | GGAA (| GAAGAAA | ACGA | GCTA | AGAT | TA S | rgato | GTGA | C A | GTGG: | rggci | r · | 840 |
| | CTCTTTGCT | G TGTGCT | GGGC | ACCATTO | CCAT | GTTC | TCCA | TA : | rga To | ATTO | A A | raca(| TAAT | r | 900 |
| | TTTGAAAAG | G AATATG | ATGA | TGTCAC | AATC | AAGA | ATGAT | TT : | TTGCT | TATCO | T G | CAAA | TATT | C | 960 |
| | GGATTTTCC | A ACTCCA | TCTG | TAATCCC | CATT | GTC1 | ATGC. | AT : | TTATO | AATG | A A | AACT: | rcaa <i>i</i> | 4 : | 1020 |
| 10 | AAAAATGTT | TGTCTG | CAGT | TTGTTAI | TGC | ATA | TAAA | TA A | AAACO | TTCI | C TO | CCAG | CACA | 4 : | 1080 |
| | AGGCATGGA | A ATTCAG | GAAT | TACAATO | ATG | CGG | AGAA | AG (| CAAAC | TTTT | C C | CTCAC | GAGAC | 3 : | 1140 |
| | AATCCAGTG | G AGGAAA | CCAA . | AGGAGAZ | AGCA | TTC | AGTGA | TG (| GCAAC | CATTO | A A | STCA | ATTO | 3 : | 1200 |
| | TGTGAACAG | A CAGAGG | AGAA | GAAAAA | CTC | AAA | GACA | TC ' | TTGCT | CTCI | T T | AGGT | CTGA | A : | 1260 |
| | CTGGCTGAG | A ATTCTC | CTTT . | AGACAGI | rggg | CATT | AA | | | | | | | : | 1296 |
| 15 | (129) INF | ORMATION | FOR | SEQ ID | NO: | 128: | | | | | | | | | |
| 20 | (i) : | SEQUENCE (A) LEN (B) TYF (C) STR (D) TOF | IGTH: PE: am RANDED | 431 ami ino aci NESS: | ino a id | acids | 3 | | | | | | | | |
| | (ii) I | MOLECULE | TYPE | : prote | ein | | | | | | | | | | |
| | (xi) | SEQUENCE | DESC | RIPTION | 1: SI | EQ II | NO: | 128 | : | | | | | | |
| | Met (| Gln Ala | Leu A | | Thr | Pro | | Gln 10 | Phe | Ser | Arg | Leu | Leu 15 | Arg | |
| 25 | Asp 1 | His Asn | Leu T | hr Arg | Glu | Gln | Phe 25 | Ile | Ala | Leu | Tyr | Arg 30 | Leu | Arg | |
| | Pro 1 | Leu Val 35 | Tyr T | hr Pro | Glu | Leu 40 | Pro | Gly | Arg | Ala | Lys 45 | Leu | Ala | Leu | |
| | Val : | Leu Thr | Gly V | al Leu | Ile | Phe | Ala | Leu | Ala | Leu | Phe | Gly | Asn | Ala | |

Leu Val Phe Tyr Val Val Thr Arg Ser Lys Ala Met Arg Thr Val Thr 65 70 75 80

I I

den com

IT

u den

A The wall the

370 375 380

Glu Thr Lys Gly Glu Ala Phe Ser Asp Gly Asn Ile Glu Val Lys Leu 385 390 395 400

Cys Glu Gln Thr Glu Glu Lys Lys Leu Lys Arg His Leu Ala Leu 5 410 415

Phe Arg Ser Glu Leu Ala Glu Asn Ser Pro Leu Asp Ser Gly His
420 425 430

- (130) INFORMATION FOR SEQ ID NO:129:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2040 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
- 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

ATGGGCAGCC CCTGGAACGG CAGCGACGGC CCCGAGGGGG CGCGGGAGCC GCCGTGGCCC

GCGCTGCCGC CTTGCGACGA GCGCCGCTGC TCGCCCTTTC CCCTGGGGGC GCTGGTGCCG

20 GTGACCGCTG TGTGCCTGTG CCTGTTCGTC GTCGGGGTGA GCGGCAACGT GGTGACCGTG

ATGCTGATCG GGCGCTACCG GGACATGCGG ACCACCA ACTTGTACCT GGGCAGCATG 240

GCCGTGTCCG ACCTACTCAT CCTGCTCGGG CTGCCGTTCG ACCTGTACCG CCTCTGGCGC 25 300

TCGCGGCCCT GGGTGTTCGG GCCGCTGCTC TGCCGCCTGT CCCTCTACGT GGGCGAGGGC

TGCACCTACG CCACGCTGCT GCACATGACC GCGCTCAGCG TCGAGCGCTA CCTGGCCATC 30 - 420

TGCCGCCCGC TCCGCGCCCG CGTCTTGGTC ACCCGGCGCC GCGTCCGCGC GCTCATCGCT

GTGCTCTGGG CCGTGGCGCT GCTCTCTGCC GGTCCCTTCT TGTTCCTGGT GGGCGTCGAG 35 - 540

CAGGACCCCG GCATCTCCGT AGTCCCGGGC CTCAATGGCA CCGCGCGGAT CGCCTCCTCG

40 CCTCTCGCCT CGTCGCCGC TCTCTGGCTC TCGCGGGCGC CACCGCCGTC CCCGCCGTCG

660

GGGCCCGAGA CCGCGGAGGC CGCGGCGCTG TTCAGCCGCG AATGCCGGCC GAGCCCCGCG

CAGCTGGGCG CGCTGCGTGT CATGCTGTGG GTCACCACCG CCTACTTCTT CCTGCCCTTT 780

CTGTGCCTCA GCATCCTCTA CGGGCTCATC GGGCGGAGC TGTGGAGCAG CCGGCGGCCG

CTGCGAGGCC CGGCCGCCTC GGGGCGGGAG AGAGGCCACC GGCAGACCAA ACGCGTCCTG

15 CGTAAGTGGA GCCGCCGTGG TTCCAAAGAC GCCTGCCTGC AGTCCGCCCC GCCGGGGACC 960

GCGCAAACGC TGGGTCCCCT TCCCCTGCTC GCCCAGCTCT GGGCGCCGCT TCCAGCTCCC 1020

20

ij.

Peril Service

Ш

W 118

[3]

٠...

1

TTTCCTATTT CGATTCCAGC CTCCACCCGC CGGTACTTCC CATCCCCCGA GAAAACCATG

TCCTGTCCCC CAGGAGCTCT GGGGGACCCC AGGGCGCTTT GAGGGTGGGA TCCCCGGATC 25 1140

CGATTCAGTA ACCAGCAGTG CTTTTCCAGA GCCTCTGAGA CCAGAAAGGA GAGTTGGTAA 1200

30 TTCTTAATCC AACCACCTGT TAGATGCCAC AAATGAGGAG TCCTCACAGT GCTCTTGAGA 1260

AGACGAGGGA GATTTCATTA AGCTAAAATT TTTTATTTAA TGTTAAGTGA TGCTGAAGGC 1320

35

50

TAAAGTAAAC CTTGCTCGTA TCAAAAAGTA AAGATTGTGC AGACCTGTTG TAGAATTCTT 1380

TTCAACAGAG AACAGAAAAC TTGTCTCCGA AGTGGGTTTG TGGAAGGAAG CCTGCCAAGG
40 1440

CGGCTTGTTC AGAGAAATTG CTCCTTCTGG TTTATGTCCA GCCTTGATAA CACATATGGG

45 AGCCTACTAT GCAGTTTTAA AGCAAGTATC CATGCAGCCT GCAGCCTGGT CATTTTTCT 1560

GGGGTGAGGA TCTGCCTAGG TAGAAGTTTT CTCTAATTTA TTTTGCTGTT ACTTGTTATT 1620

GCAGATGGTT CCTTGTCGGG GTGGGGGGTT TATTTGCTTC CCAATGCTTT TGTTAATCCC 1680

GGTGCTGTGT CTTATGTTGC AGTGGTGGTG GTTCTGGCAT TTATAATTTG CTGGTTGCCC 55 1740

TTCCACGTTG GCAGAATCAT TTACATAAAC ACGGAAGATT CGCGGATGAT GTACTTCTCT 1800

5 CAGTACTTTA ACATCGTCGC TCTGCAACTT TTCTATCTGA GCGCATCTAT CAACCCAATC 1860

CTCTACAACC TCATTTCAAA GAAGTACAGA GCGGCGGCCT TTAAACTGCT GCTCGCAAGG

10

AAGTCCAGGC CGAGAGGCTT CCACAGAAGC AGGGACACTG CGGGGGAAGT TGCAGGGGAC 1980

ACTGGAGGAG ACACGGTGGG CTACACCGAG ACAAGCGCTA ACGTGAAGAC GATGGGATAA
15 2040

- (131) INFORMATION FOR SEQ ID NO:130:
 - (i) SEQUENCE CHARACTERISTICS:

(C) STRANDEDNESS:

- (A) LENGTH: 412 amino acids
- (B) TYPE: amino acid
- 20
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

Met Gly Ser Pro Trp Asn Gly Ser Asp Gly Pro Glu Gly Ala Arg Glu
25 1 5 10 15

Pro Pro Trp Pro Ala Leu Pro Pro Cys Asp Glu Arg Arg Cys Ser Pro 20 25 30

Phe Pro Leu Gly Ala Leu Val Pro Val Thr Ala Val Cys Leu Cys Leu 35 40 45

30 Phe Val Val Gly Val Ser Gly Asn Val Val Thr Val Met Leu Ile Gly 50 60

Arg Tyr Arg Asp Met Arg Thr Thr Thr Asn Leu Tyr Leu Gly Ser Met 65 70 75 80

Ala Val Ser Asp Leu Leu Ile Leu Leu Gly Leu Pro Phe Asp Leu Tyr 85 90 95

Arg Leu Trp Arg Ser Arg Pro Trp Val Phe Gly Pro Leu Leu Cys Arg 100 105 110

Leu Ser Leu Tyr Val Gly Glu Gly Cys Thr Tyr Ala Thr Leu Leu His

40 Met Thr Ala Leu Ser Val Glu Arg Tyr Leu Ala Ile Cys Arg Pro Leu 130 135 140

| | Arg 145 | Ala | Arg | Val | Leu | Val 150 | Thr | Arg | Arg | Arg | Val 155 | Arg | Ala | Leu | Ile | Ala 160 |
|----|------------|------------|------------|------------|------------|--------------|------------|------------|------------|------------|-----------------|--------------|--------------|------------|------------|------------|
| | Val | Leu | Trp | Ala | Val 165 | Ala | Leu | Leu | Ser | Ala 170 | Gly | Pro | Phe | Leu | Phe 175 | Leu |
| 5 | Val | Gly | Val | Glu 180 | Gln | Asp | Pro | Gly | Ile 185 | Ser | Val | Val | Pro | Gly 190 | Leu | Asn |
| | Gly | Thr | Ala 195 | Arg | Ile | Ala | Ser | Ser 200 | Pro | Leu | Ala | Ser | Ser 205 | Pro | Pro | Leu |
| 10 | Trp | Leu 210 | Ser | Arg | Ala | Pro | Pro 215 | Pro | Ser | Pro | Pro | Ser 220 | Gly | Pro | Glu | Thr |
| | Ala 225 | Glu | Ala | Ala | Ala | Leu 230 | Phe | Ser | Arg | Glu | Cys 235 | Arg | Pro | Ser | Pro | Ala 240 |
| | Gln | Leu | Gly | Ala | Leu 245 | Arg | Val | Met | Leu | Trp 250 | Val | Thr | Thr | Ala | Tyr 255 | Phe |
| 15 | Phe | Leu | Pro | Phe 260 | Leu | Cys | Leu | Ser | Ile 265 | Leu | Tyr | Gly | Leu | Ile 270 | Gly | Arg |
| | Glu | Leu | Trp 275 | Ser | Ser | Arg | Arg | Pro 280 | Leu | Arg | Gly | Pro | Ala 285 | Ala | Ser | Gly |
| 20 | Arg | Glu 290 | Arg | Gly | His | Arg | Gln 295 | Thr | Lys | Arg | Val | Leu 300 | | Val | Val | Val |
| | Leu 305 | Ala | Phe | Ile | Ile | Cys 310 | Trp | Leu | Pro | Phe | His 315 | Val | Gly | Arg | Ile | Ile 320 |
| | Tyr | Ile | Asn | Thr | Glu 325 | | Ser | Arg | Met | : Met | | Phe | Ser | Gln | Tyr 335 | Phe |
| 25 | Asn | Ile | Val | Ala 340 | | Gln | . Leu | Phe | туг 345 | | . Ser | Ala | Ser | 350 | | Pro |
| | Ile | Leu | Tyr 355 | | Leu | ılle | : Ser | 1 Lys | | Tyr | Arg | , Ala | a Ala 365 | a Ala | . Phe | . Lys |
| 30 | Leu | Leu 370 | | ı Ala | . Arg | l Lys | 375 | | g Pro | Arg | g Gly | 7 Phe 380 | | s Arg | g Ser | Arg |
| | Asp 385 | | · Ala | Gly | Glı | 1 Val 390 | | a Gly | / Asj | o Thi | Gl ₃ | | y Ası | Thi | r Val | Gly 400 |
| | Tyr | Thr | Glu | ı Thr | Se: | | a Asr | n Val | L Ly | s Thi | | Gly | 7 | | | |

- 35 (132) INFORMATION FOR SEQ ID NO:131:
 - (i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 1344 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

ATGGAGCTGC TAAAGCTGAA CCGGAGCGTG CAGGGAACCG GACCCGGGCC GGGGGCTTCC

CTGTGCCGCC CGGGGGCGCC TCTCCTCAAC AGCAGCAGTG TGGGCAACCT CAGCTGCGAG

10 CCCCCTCGCA TTCGCGGAGC CGGGACACGA GAATTGGAGC TGGCCATTAG AATCACTCTT 180

TACGCAGTGA TCTTCCTGAT GAGCGTTGGA GGAAATATGC TCATCATCGT GGTCCTGGGA 240

CTGAGCCGCC GCCTGAGGAC TGTCACCAAT GCCTTCCTCC TCTCACTGGC AGTCAGCGAC
15 300

CTCCTGCTGG CTGTGGCTTG CATGCCCTTC ACCCTCCTGC CCAATCTCAT GGGCACATTC 360

ATCTTTGGCA CCGTCATCTG CAAGGCGGTT TCCTACCTCA TGGGGGTGTC TGTGAGTGTG 420

20 TCCACGCTAA GCCTCGTGGC CATCGCACTG GAGCGATATA GCGCCATCTG CCGACCACTG 480

CAGGCACGAG TGTGGCAGAC GCGCTCCCAC GCGGCTCGCG TGATTGTAGC CACGTGGCTG 540

CTGTCCGGAC TACTCATGGT GCCCTACCCC GTGTACACTG TCGTGCAACC AGTGGGGCCT
25 600

CGTGTGCTGC AGTGCGTGCA TCGCTGGCCC AGTGCGCGGG TCCGCCAGAC CTGGTCCGTA

CTGCTGCTTC TGCTCTTGTT CTTCATCCCA GGTGTGGTTA TGGCCGTGGC CTACGGGCTT 720

30 ATCTCTCGCG AGCTCTACTT AGGGCTTCGC TTTGACGGCG ACAGTGACAG CGACAGCCAA 780

AGCAGGGTCC GAAACCAAGG CGGGCTGCCA GGGGCTGTTC ACCAGAACGG GCGTTGCCGG 840

CCTGAGACTG GCGCGGTTGG CAAAGACAGC GATGGCTGCT ACGTGCAACT TCCACGTTCC 35 900

CGGCCTGCCC TGGAGCTGAC GGCGCTGACG GCTCCTGGGC CGGGATCCGG CTCCCGGCCC

ACCCAGGCCA AGCTGCTGGC TAAGAAGCGC GTGAAACGAA TGTTGCTGGT GATCGTTGTG

CTTTTTTTC TGTGTTGGTT GCCAGTTTAT AGTGCCAACA CGTGGCGCGC CTTTGATGGC

CCGGGTGCAC ACCGAGCACT CTCGGGTGCT CCTATCTCCT TCATTCACTT GCTGAGCTAC 1140

GCCTCGGCCT GTGTCAACCC CCTGGTCTAC TGCTTCATGC ACCGTCGCTT TCGCCAGGCC

10 TGCCTGGAAA CTTGCGCTCG CTGCTGCCCC CGGCCTCCAC GAGCTCGCCC CAGGGCTCTT 1260

CCCGATGAGG ACCCTCCAC TCCCTCCATT GCTTCGCTGT CCAGGCTTAG CTACACCACC 1320

ATCAGCACAC TGGGCCCTGG CTGA

15 1344

35

- (133) INFORMATION FOR SEQ ID NO:132:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 447 amino acids
 - (B) TYPE: amino acid
- (C) STRANDEDNESS: 20
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

Met Glu Leu Leu Lys Leu Asn Arg Ser Val Gln Gly Thr Gly Pro Gly 25 5

> Pro Gly Ala Ser Leu Cys Arg Pro Gly Ala Pro Leu Leu Asn Ser Ser 25

> Ser Val Gly Asn Leu Ser Cys Glu Pro Pro Arg Ile Arg Gly Ala Gly 35

Thr Arg Glu Leu Glu Leu Ala Ile Arg Ile Thr Leu Tyr Ala Val Ile 30

> Phe Leu Met Ser Val Gly Gly Asn Met Leu Ile Ile Val Val Leu Gly 70

Leu Ser Arg Arg Leu Arg Thr Val Thr Asn Ala Phe Leu Leu Ser Leu 85

Ala Val Ser Asp Leu Leu Leu Ala Val Ala Cys Met Pro Phe Thr Leu

| | | | | 100 | | | | | 105 | | | | | 110 | | |
|----|------------|--------------|------------|------------|------------|------------|------------|------------|--------------|------------|--------------|------------|------------|------------|------------|--------------|
| | Leu | Pro | Asn 115 | Leu | Met | Gly | Thr | Phe 120 | Ile | Phe | Gly | Thr | Val 125 | Ile | Cys | Lys |
| 5 | Ala | Val 130 | Ser | Tyr | Leu | Met | Gly 135 | Val | Ser | Val | | Val 140 | Ser | Thr | Leu | Ser |
| | Leu 145 | Val | Ala | Ile | Ala | Leu 150 | Glu | Arg | Tyr | Ser | Ala 155 | Ile | Cys | Arg | Pro | Leu 160 |
| | Gln | Ala | Arg | Val | Trp 165 | Gln | Thr | Arg | Ser | His 170 | Ala | Ala | Arg | Val | Ile 175 | Val |
| 10 | Ala | Thr | Trp | Leu 180 | Leu | Ser | Gly | Leu | Leu 185 | Met | Val | Pro | Tyr | Pro 190 | Val | Tyr |
| | Thr | Val | Val 195 | Gln | Pro | Val | Gly | Pro 200 | Arg | Val | Leu | Gln | Cys 205 | Val | His | Arg |
| 15 | Trp | Pro 210 | Ser | Ala | Arg | Val | Arg 215 | Gln | Thr | Trp | Ser | Val 220 | Leu | Leu | Leu | Leu |
| | Leu 225 | | Phe | Phe | Ile | Pro 230 | Gly | Val | Val | Met | Ala 235 | Val | Ala | Tyr | Gly | Leu 240 |
| | Ile | Ser | Arg | Glu | Leu 245 | | Leu | Gly | Leu | Arg 250 | Phe | Asp | Gly | Asp | Ser 255 | Asp |
| 20 | Ser | Asp | Ser | Gln 260 | | Arg | Val | Arg | Asn 265 | | Gly | Gly | Leu | Pro 270 | Gly | Ala |
| | Val | His | Gln 275 | Asn | Gly | Arg | Cys | Arg 280 | | Glu | Thr | Gly | Ala 285 | Val | Gly | Lys |
| 25 | Asp | Ser 290 | | Gly | Cys | Tyr | Val 295 | | Leu | Pro | Arg | Ser 300 | | Pro | Ala | Leu |
| | Glu 305 | | Thr | Ala | Leu | Thr 310 | | Pro | Gly | Pro | Gly 315 | | Gly | Ser | Arg | Pro 320 |
| | Thr | Gln | Ala | Lys | Leu 325 | | Ala | Lys | . Lys | 330 | | Lys | arg | Met | Leu 335 | Leu |
| 30 | Va] | l Ile | e Val | . Val | | ı Phe | Phe | e Lev | 1 Cys 345 | | Leu | Pro | o Val | Туг 350 | | Ala |
| | Ası | n Thi | Trp 355 | | g Ala | a Phe | e Asp | 360 | | Gly | / Ala | Hi: | 365 | | ı Leı | ser |
| 35 | Va: | 1 Ala 370 | | o Ile | e Sei | c Phe | 375 | | s Lei | ı Lei | ı Ser | ту: 38 | | a Sei | ala | a Cys |
| | Va: | | n Pro |) Lei | ı Val | 1 Ty: | | s Phe | e Me | t Hi | s Arg 395 | | g Phe | e Arg | g Gli | n Ala 400 |

20

25

30

Cys Leu Glu Thr Cys Ala Arg Cys Cys Pro Arg Pro Pro Arg Ala Arg

Pro Arg Ala Leu Pro Asp Glu Asp Pro Pro Thr Pro Ser Ile Ala Ser 420 425 430

5 Leu Ser Arg Leu Ser Tyr Thr Thr Ile Ser Thr Leu Gly Pro Gly
435 440 445

(134) INFORMATION FOR SEQ ID NO:133:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1014 base pairs
- 10 (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

| ATGAACAGCA | CATGTATTGA | AGAACAGCAT | GACCTGGATC | ACTATTTGTT | TCCCATTGTT | 60 |
|------------|------------|------------|------------|------------|------------|------|
| TACATCTTTG | TGATTATAGT | CAGCATTCCA | GCCAATATTG | GATCTCTGTG | TGTGTCTTTC | 120 |
| CTGCAAGCAA | AGAAGGAAAG | TGAACTAGGA | ATTTACCTCT | TCAGTTTGTC | ACTATCAGAT | 180 |
| TTACTCTATG | CATTAACTCT | CCCTTTATGG | ATTGATTATA | CTTGGAATAA | AGACAACTGG | 240 |
| ACTTTCTCTC | CTGCCTTGTG | CAAAGGGAGT | GCTTTTCTCA | TGTACATGAA | TTTTTACAGC | 300 |
| AGCACAGCAT | TCCTCACCTG | CATTGCCGTT | GATCGGTATT | TGGCTGTTGT | CTACCCTTTG | 360 |
| AAGTTTTTT | TCCTAAGGAC | AAGAAGATTT | GCACTCATGG | TCAGCCTGTC | CATCTGGATA | 420 |
| TTGGAAACCA | TCTTCAATGC | TGTCATGTTG | TGGGAAGATG | AAACAGTTGT | TGAATATTGC | 480 |
| GATGCCGAAA | AGTCTAATTT | TACTTTATGC | TATGACAAAT | ACCCTTTAGA | GAAATGGCAA | 540 |
| ATCAACCTCA | ACTTGTTCAG | GACGTGTACA | GGCTATGCAA | TACCTTTGGT | CACCATCCTG | 600 |
| ATCTGTAACC | GGAAAGTCTA | CCAAGCTGTG | CGGCACAATA | AAGCCACGGA | AAACAAGGAA | 660 |
| AAGAAGAGAA | TCAAAAAACT | ACTTGTCAGC | ATCACAGTTA | CTTTTGTCTT | ATGCTTTACT | 720 |
| CCCTTTCATG | TGATGTTGCT | GATTCGCTGC | ATTTTAGAGC | ATGCTGTGAA | CTTCGAAGAC | 780 |
| CACAGCAATT | CTGGGAAGCG | AACTTACACA | ATGTATAGAA | TCACGGTTGC | ATTAACAAGT | 840 |
| TTAAATTGTG | TTGCTGATCC | AATTCTGTAC | TGTTTTGTTA | CCGAAACAGG | AAGATATGAT | 900 |
| ATGTGGAATA | TATTAAAATT | CTGCACTGGG | AGGTGTAATA | CATCACAAAG | ACAAAGAAAA | 960 |
| CGCATACTTT | CTGTGTCTAC | AAAAGATACT | ATGGAATTAG | AGGTCCTTGA | GTAG | 1014 |

- (135) INFORMATION FOR SEQ ID NO:134:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 337 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:
- Met Asn Ser Thr Cys Ile Glu Glu Gln His Asp Leu Asp His Tyr Leu 10 1 5 10 15
 - Phe Pro Ile Val Tyr Ile Phe Val Ile Ile Val Ser Ile Pro Ala Asn 20 25 30
 - Ile Gly Ser Leu Cys Val Ser Phe Leu Gln Ala Lys Lys Glu Ser Glu
 35 40 45
- Leu Gly Ile Tyr Leu Phe Ser Leu Ser Leu Ser Asp Leu Leu Tyr Ala
 50 55 60
 - Leu Thr Leu Pro Leu Trp Ile Asp Tyr Thr Trp Asn Lys Asp Asn Trp 65 70 75 80
- Thr Phe Ser Pro Ala Leu Cys Lys Gly Ser Ala Phe Leu Met Tyr Met 20 85 90 95
 - Asn Phe Tyr Ser Ser Thr Ala Phe Leu Thr Cys Ile Ala Val Asp Arg
 - Tyr Leu Ala Val Val Tyr Pro Leu Lys Phe Phe Phe Leu Arg Thr Arg
- 25 Arg Phe Ala Leu Met Val Ser Leu Ser Ile Trp Ile Leu Glu Thr Ile 130 135 140
 - Phe Asn Ala Val Met Leu Trp Glu Asp Glu Thr Val Val Glu Tyr Cys 145 150 150
- Asp Ala Glu Lys Ser Asn Phe Thr Leu Cys Tyr Asp Lys Tyr Pro Leu 30 165 170 175
 - Glu Lys Trp Gln Ile Asn Leu Asn Leu Phe Arg Thr Cys Thr Gly Tyr 180 185 190
 - Ala Ile Pro Leu Val Thr Ile Leu Ile Cys Asn Arg Lys Val Tyr Gln 195 200 205
- Ala Val Arg His Asn Lys Ala Thr Glu Asn Lys Glu Lys Lys Arg Ile
 210 215 220

Lys Lys Leu Leu Val Ser Ile Thr Val Thr Phe Val Leu Cys Phe Thr 225 230 235 235

Pro Phe His Val Met Leu Leu Ile Arg Cys Ile Leu Glu His Ala Val 245 250 255

Asn Phe Glu Asp His Ser Asn Ser Gly Lys Arg Thr Tyr Thr Met Tyr 260 265 270

Arg Ile Thr Val Ala Leu Thr Ser Leu Asn Cys Val Ala Asp Pro Ile 275 280 285

Leu Tyr Cys Phe Val Thr Glu Thr Gly Arg Tyr Asp Met Trp Asn Ile 10 290 295 300

Leu Lys Phe Cys Thr Gly Arg Cys Asn Thr Ser Gln Arg Gln Arg Lys 305 310 315 315

Arg Ile Leu Ser Val Ser Thr Lys Asp Thr Met Glu Leu Glu Val Leu 325 330 335

15 Glu

20

(136) INFORMATION FOR SEQ ID NO:135:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 999 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:
- 25 ATGGTGAACT CCACCCACCG TGGGATGCAC ACTTCTCTGC ACCTCTGGAA CCGCAGCAGT

TACAGACTGC ACAGCAATGC CAGTGAGTCC CTTGGAAAAG GCTACTCTGA TGGAGGGTGC 120

TACGAGCAAC TTTTTGTCTC TCCTGAGGTG TTTGTGACTC TGGGTGTCAT CAGCTTGTTG

GAGAATATCT TAGTGATTGT GGCAATAGCC AAGAACAAGA ATCTGCATTC ACCCATGTAC 240

TTTTTCATCT GCAGCTTGGC TGTGGCTGAT ATGCTGGTGA GCGTTTCAAA TGGATCAGAA 300

35 ACCATTATCA TCACCCTATT AAACAGTACA GATACGGATG CACAGAGTTT CACAGTGAAT 360

1 i II

1 42 H H

IT H

12

(T M. C.

25

ATTGATAATG TCATTGACTC GGTGATCTGT AGCTCCTTGC TTGCATCCAT TTGCAGCCTG

CTTTCAATTG CAGTGGACAG GTACTTTACT ATCTTCTATG CTCTCCAGTA CCATAACATT

5 ATGACAGTTA AGCGGGTTGG GATCAGCATA AGTTGTATCT GGGCAGCTTG CACGGTTTCA

GGCATTTTGT TCATCATTTA CTCAGATAGT AGTGCTGTCA TCATCTGCCT CATCACCATG

TTCTTCACCA TGCTGGCTCT CATGGCTTCT CTCTATGTCC ACATGTTCCT GATGGCCAGG 10 660

CTTCACATTA AGAGGATTGC TGTCCTCCCC GGCACTGGTG CCATCCGCCA AGGTGCCAAT

ATGAAGGGAA AAATTACCTT GACCATCCTG ATTGGCGTCT TTGTTGTCTG CTGGGCCCCA 780

15 TTCTTCCTCC ACTTAATATT CTACATCTCT TGTCCTCAGA ATCCATATTG TGTGTGCTTC

ATGTCTCACT TTAACTTGTA TCTCATACTG ATCATGTGTA ATTCAATCAT CGATCCTCTG

ATTTATGCAC TCCGGAGTCA AGAACTGAGG AAAACCTTCA AAGAGATCAT CTGTTGCTAT 20 960

CCCCTGGGAG GCCTTTGTGA CTTGTCTAGC AGATATTAA 999

- (137) INFORMATION FOR SEQ ID NO:136:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 332 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
- 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

Met Val Asn Ser Thr His Arg Gly Met His Thr Ser Leu His Leu Trp 5 10

Asn Arg Ser Ser Tyr Arg Leu His Ser Asn Ala Ser Glu Ser Leu Gly

Lys Gly Tyr Ser Asp Gly Gly Cys Tyr Glu Gln Leu Phe Val Ser Pro 35 40

| | Glu | Val 50 | Phe | Val | Thr | Leu | Gly 55 | Val | Ile | Ser | Leu | Leu 60 | Glu | Asn | Ile | Leu |
|----|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | Val 65 | Ile | Val | Ala | Ile | Ala 70 | Lys | Asn | Lys | Asn | Leu 75 | His | Ser | Pro | Met | Tyr 80 |
| 5 | Phe | Phe | Ile | Cys | Ser 85 | Leu | Ala | Val | Ala | Asp 90 | Met | Leu | Val | Ser | Val 95 | Ser |
| | Asn | Gly | Ser | Glu 100 | Thr | Ile | Ile | Ile | Thr 105 | Leu | Leu | Asn | Ser | Thr 110 | Asp | Thr |
| 10 | Asp | Ala | Gln 115 | Ser | Phe | Thr | Val | Asn 120 | Ile | Asp | Asn | Val | Ile 125 | Asp | Ser | Val |
| | Ile | Cys 130 | Ser | Ser | Leu | Leu | Ala 135 | Ser | Ile | Cys | Ser | Leu 140 | Leu | Ser | Ile | Ala |
| | Val 145 | Asp | Arg | Tyr | Phe | Thr 150 | Ile | Phe | Tyr | Ala | Leu 155 | Gln | Tyr | His | Asn | Ile 160 |
| 15 | Met | Thr | Val | Lys | Arg 165 | Val | Gly | Ile | Ser | Ile 170 | Ser | Cys | Ile | Trp | Ala 175 | Ala |
| | Cys | Thr | Val | Ser 180 | Gly | Ile | Leu | Phe | Ile 185 | Ile | Tyr | Ser | Asp | Ser 190 | Ser | Ala |
| 20 | Val | Ile | Ile 195 | Cys | Leu | Ile | Thr | Met 200 | Phe | Phe | Thr | Met | Leu 205 | Ala | Leu | Met |
| | Ala | Ser 210 | Leu | Tyr | Val | His | Met 215 | Phe | Leu | Met | Ala | Arg 220 | Leu | His | Ile | Lys |
| | Arg 225 | Ile | Ala | Val | Leu | Pro 230 | Gly | Thr | Gly | Ala | Ile 235 | Arg | Gln | Gly | Ala | Asn 240 |
| 25 | Met | Lys | Gly | Lys | Ile 245 | Thr | Leu | Thr | Ile | Leu 250 | Ile | Gly | Val | Phe | Val 255 | Val |
| | Cys | Trp | Ala | Pro 260 | Phe | Phe | Leu | His | Leu 265 | Ile | Phe | Tyr | Ile | Ser 270 | Cys | Pro |
| 30 | Gln | Asn | Pro 275 | Tyr | Cys | Val | Cys | Phe 280 | Met | Ser | His | Phe | Asn 285 | Leu | Tyr | Leu |
| | Ile | Leu 290 | Ile | Met | Cys | Asn | Ser 295 | Ile | Ile | Asp | Pro | Leu 300 | Ile | Tyr | Ala | Leu |
| | Arg 305 | | Gln | Glu | Leu | Arg 310 | Lys | Thr | Phe | Lys | Glu 315 | Ile | Ile | Cys | Суз | Tyr 320 |

35 Pro Leu Gly Gly Leu Cys Asp Leu Ser Ser Arg Tyr 325 330

(138) INFORMATION FOR SEQ ID NO:137:

The first first that was the control of the control

| 5 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
|----|--|-----|
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:137: | |
| | GCCAATATGA AGGGAAAAAT TACCTTGACC ATC 33 | |
| 10 | (137) INFORMATION FOR SEQ ID NO:138: | |
| 15 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| 15 | (ii) MOLECULE TYPE: DNA (genomic) | |
| | (II) MODECODE TIPE: DNA (Genomic) | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:138: | |
| | CTCCTTCGGT CCTCCTATCG TTGTCAGAAG T | |
| 20 | (140) INFORMATION FOR SEQ ID NO:139: | |
| 25 | (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1842 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| | (ii) MOLECULE TYPE: DNA (genomic) | |
| | | |
| | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:139: | |
| | ATGGGGCCCA CCCTAGCGGT TCCCACCCC TATGGCTGTA TTGGCTGTAA GCTACCCCAG | 60 |
| | CCAGAATACC CACCGGCTCT AATCATCTTT ATGTTCTGCG CGATGGTTAT CACCATCGTT | 120 |
| 30 | GTAGACCTAA TCGGCAACTC CATGGTCATT TTGGCTGTGA CGAAGAACAA GAAGCTCCGG | 180 |
| | AATTCTGGCA ACATCTTCGT GGTCAGTCTC TCTGTGGCCG ATATGCTGGT GGCCATCTAC | 240 |
| | CCATACCCTT TGATGCTGCA TGCCATGTCC ATTGGGGGCT GGGATCTGAG CCAGTTACAG | 300 |
| | TGCCAGATGG TCGGGTTCAT CACAGGGCTG AGTGTGGTCG GCTCCATCTT CAACATCGTG | 360 |

AREN-0054 - 169 - PATENT

| | GCAATCGCTA | TCAACCGTTA | CTGCTACATC | TGCCACAGCC | TCCAGTACGA | ACGGATCTTC | 420 |
|----|------------|------------|------------|--------------|--------------|------------|------|
| | AGTGTGCGCA | ATACCTGCAT | CTACCTGGTC | ATCACCTGGA | TCATGACCGT | CCTGGCTGTC | 480 |
| | CTGCCCAACA | TGTACATTGG | CACCATCGAG | TACGATCCTC | GCACCTACAC | CTGCATCTTC | 540 |
| | AACTATCTGA | ACAACCCTGT | CTTCACTGTT | ACCATCGTCT | GCATCCACTT | CGTCCTCCCT | 600 |
| 5 | CTCCTCATCG | TGGGTTTCTG | CTACGTGAGG | ATCTGGACCA | AAGTGCTGGC | GGCCCGTGAC | 660 |
| | CCTGCAGGGC | AGAATCCTGA | CAACCAACTT | GCTGAGGTTC | GCAATTTTCT | AACCATGTTT | 720 |
| | GTGATCTTCC | TCCTCTTTGC | AGTGTGCTGG | TGCCCTATCA | ACGTGCTCAC | TGTCTTGGTG | 780 |
| | GCTGTCAGTC | CGAAGGAGAT | GGCAGGCAAG | ATCCCCAACT | GGCTTTATCT | TGCAGCCTAC | 840 |
| | TTCATAGCCT | ACTTCAACAG | CTGCCTCAAC | GCTGTGATCT | ACGGGCTCCT | CAATGAGAAT | 900 |
| 10 | TTCCGAAGAG | AATACTGGAC | CATCTTCCAT | GCTATGCGGC | ACCCTATCAT | ATTCTTCCCT | 960 |
| | GGCCTCATCA | GTGATATTCG | TGAGATGCAG | GAGGCCCGTA | CCCTGGCCCG | CGCCCGTGCC | 1020 |
| | CATGCTCGCG | ACCAAGCTCG | TGAACAAGAC | CGTGCCCATG | CCTGTCCTGC | TGTGGAGGAA | 1080 |
| | ACCCCGATGA | ATGTCCGGAA | TGTTCCATTA | CCTGGTGATG | CTGCAGCTGG | CCACCCCGAC | 1140 |
| | CGTGCCTCTG | GCCACCCTAA | GCCCCATTCC | AGATCCTCCT | CTGCCTATCG | CAAATCTGCC | 1200 |
| 15 | TCTACCCACC | ACAAGTCTGT | CTTTAGCCAC | TCCAAGGCTG | CCTCTGGTCA | CCTCAAGCCT | 1260 |
| | GTCTCTGGCC | ACTCCAAGCC | TGCCTCTGGT | CACCCCAAGT | CTGCCACTGT | CTACCCTAAG | 1320 |
| | CCTGCCTCTG | TCCATTTCAA | GGGTGACTCT | GTCCATTTCA | AGGGTGACTC | TGTCCATTTC | 1380 |
| | AAGCCTGACT | CTGTTCATTT | CAAGCCTGCT | TCCAGCAACC | CCAAGCCCAT | CACTGGCCAC | 1440 |
| | CATGTCTCTG | CTGGCAGCCA | CTCCAAGTCI | GCCTTCAGTG | CTGCCACCAG | CCACCCTAAA | 1500 |
| 20 | CCCATCAAGC | CAGCTACCAG | CCATGCTGAG | CCCACCACTG | CTGACTATCC | CAAGCCTGCC | 1560 |
| | ACTACCAGCC | ACCCTAAGCC | CGCTGCTGCT | GACAACCCTG | AGCTCTCTGC | CTCCCATTGC | 1620 |
| | CCCGAGATCC | CTGCCATTGC | CCACCCTGTG | TCTGACGACA | GTGACCTCCC | TGAGTCGGCC | 1680 |
| | TCTAGCCCTG | CCGCTGGGCC | CACCAAGCCI | GCTGCCAGCC | : AGCTGGAGTC | TGACACCATC | 1740 |
| | GCTGACCTTC | CTGACCCTAC | TGTAGTCACT | r accagtacca | ATGATTACCA | TGATGTCGTG | 1800 |
| 25 | GTTGTTGATG | TTGAAGATGA | TCCTGATGA | A ATGGCTGTGT | GA | | 1842 |

(141) INFORMATION FOR SEQ ID NO:140:

⁽i) SEQUENCE CHARACTERISTICS:

⁽A) LENGTH: 613 amino acids

⁽B) TYPE: amino acid

- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:
- 5 Met Gly Pro Thr Leu Ala Val Pro Thr Pro Tyr Gly Cys Ile Gly Cys
 1 10 15
 - Lys Leu Pro Gln Pro Glu Tyr Pro Pro Ala Leu Ile Ile Phe Met Phe 20 25 30
- Cys Ala Met Val Ile Thr Ile Val Val Asp Leu Ile Gly Asn Ser Met
 10 35 40 45
 - Val Ile Leu Ala Val Thr Lys Asn Lys Lys Leu Arg Asn Ser Gly Asn 50 55 60
 - Ile Phe Val Val Ser Leu Ser Val Ala Asp Met Leu Val Ala Ile Tyr 65 70 75 80
- pro Tyr Pro Leu Met Leu His Ala Met Ser Ile Gly Gly Trp Asp Leu 85 90 95
 - Ser Gln Leu Gln Cys Gln Met Val Gly Phe Ile Thr Gly Leu Ser Val
- Val Gly Ser Ile Phe Asn Ile Val Ala Ile Ala Ile Asn Arg Tyr Cys 20 115 120 125
 - Tyr Ile Cys His Ser Leu Gln Tyr Glu Arg Ile Phe Ser Val Arg Asn 130 135 140
 - Thr Cys Ile Tyr Leu Val Ile Thr Trp Ile Met Thr Val Leu Ala Val 145 150 155 160
- 25 Leu Pro Asn Met Tyr Ile Gly Thr Ile Glu Tyr Asp Pro Arg Thr Tyr 165 170 175
 - Thr Cys Ile Phe Asn Tyr Leu Asn Asn Pro Val Phe Thr Val Thr Ile 180 185 190
- Val Cys Ile His Phe Val Leu Pro Leu Leu Ile Val Gly Phe Cys Tyr 30 195 200 205
 - Val Arg Ile Trp Thr Lys Val Leu Ala Ala Arg Asp Pro Ala Gly Gln 210 215 220
 - Asn Pro Asp Asn Gln Leu Ala Glu Val Arg Asn Phe Leu Thr Met Phe 225 230 235 240
- Val Ile Phe Leu Leu Phe Ala Val Cys Trp Cys Pro Ile Asn Val Leu 245 250 255

Thr Val Leu Val Ala Val Ser Pro Lys Glu Met Ala Gly Lys Ile Pro 260 265 270

Asn Trp Leu Tyr Leu Ala Ala Tyr Phe Ile Ala Tyr Phe Asn Ser Cys
275
280
285

535

Ala Ile Ala His Pro Val Ser Asp Asp Ser Asp Leu Pro Glu Ser Ala

560 555 545 550 Ser Ser Pro Ala Ala Gly Pro Thr Lys Pro Ala Ala Ser Gln Leu Glu 570 565 Ser Asp Thr Ile Ala Asp Leu Pro Asp Pro Thr Val Val Thr Thr Ser 585 5 Thr Asn Asp Tyr His Asp Val Val Val Asp Val Glu Asp Asp Pro 600 Asp Glu Met Ala Val

10 (142) INFORMATION FOR SEQ ID NO:141:

610

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1842 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

| | ATGGGGCCCA | CCCTAGCGGT | TCCCACCCCC | TATGGCTGTA | TTGGCTGTAA | GCTACCCCAG | 60 |
|----|------------|------------|------------|------------|------------|------------|-----|
| | CCAGAATACC | CACCGGCTCT | AATCATCTTT | ATGTTCTGCG | CGATGGTTAT | CACCATCGTT | 120 |
| 20 | GTAGACCTAA | TCGGCAACTC | CATGGTCATT | TTGGCTGTGA | CGAAGAACAA | GAAGCTCCGG | 180 |
| | AATTCTGGCA | ACATCTTCGT | GGTCAGTCTC | TCTGTGGCCG | ATATGCTGGT | GGCCATCTAC | 240 |
| | CCATACCCTT | TGATGCTGCA | TGCCATGTCC | ATTGGGGGCT | GGGATCTGAG | CCAGTTACAG | 300 |
| | TGCCAGATGG | TCGGGTTCAT | CACAGGGCTG | AGTGTGGTCG | GCTCCATCTT | CAACATCGTG | 360 |
| | GCAATCGCTA | TCAACCGTTA | CTGCTACATC | TGCCACAGCC | TCCAGTACGA | ACGGATCTTC | 420 |
| 25 | AGTGTGCGCA | ATACCTGCAT | CTACCTGGTC | ATCACCTGGA | TCATGACCGT | CCTGGCTGTC | 480 |
| | CTGCCCAACA | TGTACATTGG | CACCATCGAG | TACGATCCTC | GCACCTACAC | CTGCATCTTC | 540 |
| | AACTATCTGA | ACAACCCTGT | CTTCACTGTT | ACCATCGTCT | GCATCCACTT | CGTCCTCCCT | 600 |
| | CTCCTCATCG | TGGGTTTCTG | CTACGTGAGG | ATCTGGACCA | AAGTGCTGGC | GGCCCGTGAC | 660 |
| | CCTGCAGGGC | AGAATCCTGA | CAACCAACTT | GCTGAGGTTC | GCAATAAACT | AACCATGTTT | 720 |
| 30 | GTGATCTTCC | TCCTCTTTGC | AGTGTGCTGG | TGCCCTATCA | ACGTGCTCAC | TGTCTTGGTG | 780 |
| | GCTGTCAGTC | CGAAGGAGAT | GGCAGGCAAG | ATCCCCAACT | GGCTTTATCT | TGCAGCCTAC | 840 |

| | | | | | 3 GGGGGGGG | C | 900 |
|----|------------|------------|-------------|--------------|-------------|--------------|------|
| | TTCATAGCCT | ACTTCAACAG | CTGCCTCAAC | GCTGTGATCT | ACGGGCTCCT | CAATGAGAAT | 300 |
| • | TTCCGAAGAG | AATACTGGAC | CATCTTCCAT | GCTATGCGGC | ACCCTATCAT | ATTCTTCTCT | 960 |
| | GGCCTCATCA | GTGATATTCG | TGAGATGCAG | GAGGCCCGTA | CCCTGGCCCG | CGCCCGTGCC | 1020 |
| | CATGCTCGCG | ACCAAGCTCG | TGAACAAGAC | CGTGCCCATG | CCTGTCCTGC | TGTGGAGGAA | 1080 |
| 5 | ACCCCGATGA | ATGTCCGGAA | TGTTCCATTA | CCTGGTGATG | CTGCAGCTGG | CCACCCCGAC | 1140 |
| | CGTGCCTCTG | GCCACCCTAA | GCCCCATTCC | AGATCCTCCT | CTGCCTATCG | CAAATCTGCC | 1200 |
| | TCTACCCACC | ACAAGTCTGT | CTTTAGCCAC | TCCAAGGCTG | CCTCTGGTCA | CCTCAAGCCT | 1260 |
| | GTCTCTGGCC | ACTCCAAGCC | TGCCTCTGGT | CACCCCAAGT | CTGCCACTGT | CTACCCTAAG | 1320 |
| | CCTGCCTCTG | TCCATTTCAA | GGCTGACTCT | GTCCATTTCA | AGGGTGACTC | TGTCCATTTC | 1380 |
| 10 | AAGCCTGACT | CTGTTCATTT | CAAGCCTGCT | TCCAGCAACC | CCAAGCCCAT | CACTGGCCAC | 1440 |
| | CATGTCTCTG | CTGGCAGCCA | CTCCAAGTCT | GCCTTCAATG | CTGCCACCAG | CCACCCTAAA | 1500 |
| | CCCATCAAGC | CAGCTACCAG | CCATGCTGAG | CCCACCACTG | CTGACTATCC | CAAGCCTGCC | 1560 |
| | ACTACCAGCC | ACCCTAAGCC | CGCTGCTGCT | GACAACCCTG | AGCTCTCTGC | CTCCCATTGC | 1620 |
| | CCCGAGATCC | CTGCCATTGC | CCACCCTGTG | TCTGACGACA | GTGACCTCC | TGAGTCGGCC | 1680 |
| 15 | TCTAGCCCTG | CCGCTGGGCC | CACCAAGCCI | GCTGCCAGCC | AGCTGGAGT | TGACACCATC | 1740 |
| | GCTGACCTTC | CTGACCCTAC | TGTAGTCACI | ACCAGTACCA | A ATGATTACC | A TGATGTCGTG | 1800 |
| | GTTGTTGATG | TTGAAGATGA | A TCCTGATGA | A ATGGCTGTGT | r ga | | 1842 |

- (143) INFORMATION FOR SEQ ID NO:142:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 613 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: protein
- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

Met Gly Pro Thr Leu Ala Val Pro Thr Pro Tyr Gly Cys Ile Gly Cys 1 5 10 15

Lys Leu Pro Gln Pro Glu Tyr Pro Pro Ala Leu Ile Ile Phe Met Phe 20 25 30

Cys Ala Met Val Ile Thr Ile Val Val Asp Leu Ile Gly Asn Ser Met 35 40 45

| | Val | Ile 50 | Leu | Ala | Val | Thr | Lys 55 | Asn | Lys | Lys | Leu | Arg 60 | Asn | Ser | Gly | Asn |
|----|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | Ile 65 | Phe | Val | Val | Ser | Leu 70 | Ser | Val | Ala | Asp | Met 75 | Leu | Val | Ala | Ile | Tyr 80 |
| 5 | Pro | Tyr | Pro | Leu | Met 85 | Leu | His | Ala | Met | Ser 90 | Ile | Gly | Gly | Trp | Asp 95 | Leu |
| | Ser | Gln | Leu | Gln 100 | Cys | Gln | Met | Val | Gly 105 | Phe | Ile | Thr | Gly | Leu 110 | Ser | Val |
| 10 | Val | Gly | Ser 115 | Ile | Phe | Asn | Ile | Val 120 | Ala | Ile | Ala | Ile | Asn 125 | Arg | Tyr | Cys |
| | Tyr | Ile 130 | Cys | His | Ser | Leu | Gln 135 | Tyr | Glu | Arg | Ile | Phe 140 | Ser | Val | Arg | Asn |
| | Thr 145 | Cys | Ile | Tyr | Leu | Val 150 | Ile | Thr | Trp | Ile | Met 155 | Thr | Val | Leu | Ala | Val 160 |
| 15 | Leu | Pro | Asn | Met | Tyr 165 | Ile | Gly | Thr | Ile | Glu 170 | Tyr | Asp | Pro | Arg | Thr 175 | Tyr |
| | Thr | Cys | Ile | Phe 180 | Asn | Tyr | Leu | Asn | Asn 185 | Pro | Val | Phe | Thr | Val 190 | Thr | Ile |
| 20 | Val | Cys | Ile 195 | His | Phe | Val | Leu | Pro 200 | | Leu | Ile | Val | Gly 205 | | Cys | Tyr |
| | Val | Arg 210 | | Trp | Thr | Lys | Val 215 | | Ala | Ala | Arg | 220 | | Ala | Gly | Gln |
| | Asn 225 | | Asp | Asn | Gln | Leu 230 | | Glu | Val | Arg | Asn 235 | | Leu | Thr | Met | Phe 240 |
| 25 | Val | Ile | Phe | Leu | Leu 245 | | Ala | . Val | . Cys | 250 | | Pro | Ile | . Asn | Val 255 | Leu |
| | Thr | Val | . Leu | Val 260 | | . Val | . Ser | Pro | 265 | | ı Met | : Ala | a Gly | 270 | | Pro |
| 30 | Asr | ı Trp | Leu 275 | | Leu | ı Ala | a Ala | Туг 280 | | e Ile | e Ala | а Туі | 285 | | Ser | Cys |
| | Leu | 290 | | . Val | . Ile | э Туг | Gly 295 | | ı Leı | ı Ası | ı Glı | 300 | | e Arg | J Arc | g Glu |
| | Ту: 305 | | Thi | : Ile | Phe | His 310 | | a Met | arg | g His | 319 | | e Ile | e Phe | e Phe | 320 |
| 35 | Gl | / Lei | ı Ile | e Sei | 325 | | e Arg | g Glı | ı Met | 33 | | u Ala | a Arg | g Thi | 33! | ı Ala |
| | Arg | g Ala | a Arg | g Ala | a His | s Ala | a Arç | g As | p Gli | n Al | a Ar | g Gl | u Gl | n Ası | Arg | g Ala |

| | | 340 | | 345 | | 350 |
|----|------------------|--------------------|------------------|-----------------------|---------------------|--------------------|
| | | ys Pro Ala 55 | | Glu Thr Pro Me 360 | et Asn Val . 365 | Arg Asn Val |
| 5 | Pro Leu Pro 370 | ro Gly Asp | Ala Ala A 375 | Ala Gly His P | ro Asp Arg | Ala Ser Gly |
| | His Pro L | ys Pro His | Ser Arg S | Ser Ser Ser A | la Tyr Arg 95 | Lys Ser Ala 400 |
| | Ser Thr H | is His Lys 405 | Ser Val I | Phe Ser His Se 410 | er Lys Ala | Ala Ser Gly 415 |
| 10 | His Leu L | ys Pro Val 420 | Ser Gly H | His Ser Lys P 425 | ro Ala Ser | Gly His Pro 430 |
| | - | la Thr Val 35 | | Lys Pro Ala S 440 | er Val His 445 | Phe Lys Ala |
| 15 | Asp Ser V 450 | al His Phe | Lys Gly 455 | Asp Ser Val H | is Phe Lys 460 | Pro Asp Ser |
| | Val His P 465 | Phe Lys Pro | Ala Ser 470 | Ser Asn Pro L 4 | ys Pro Ile 75 | Thr Gly His 480 |
| | His Val S | Ser Ala Gly 485 | | Ser Lys Ser A 490 | ala Phe Asn | Ala Ala Thr 495 |
| 20 | Ser His F | Pro Lys Pro 500 | lle Lys | Pro Ala Thr S 505 | Ser His Ala | Glu Pro Thr 510 |
| | | Asp Tyr Pro | Lys Pro | Ala Thr Thr S 520 | Ser His Pro 525 | Lys Pro Ala |
| 25 | Ala Ala A 530 | Asp Asn Pro | Glu Leu 535 | Ser Ala Ser F | His Cys Pro 540 | Glu Ile Pro |
| | Ala Ile A 545 | Ala His Pro | Val Ser 550 | Asp Asp Ser A | Asp Leu Pro 555 | Glu Ser Ala 560 |
| | Ser Ser | Pro Ala Ala 56! | | Thr Lys Pro 2 570 | Ala Ala Ser | Gln Leu Glu 575 |
| 30 | Ser Asp ' | Thr Ile Ala 580 | a Asp Leu | Pro Asp Pro ' | Thr Val Val | Thr Thr Ser |
| | | Asp Tyr Hi 595 | s Asp Val | Val Val Val 600 | Asp Val Glu 605 | |
| 35 | Asp Glu 610 | Met Ala Va | 1 | | | |

(144) INFORMATION FOR SEQ ID NO:143:

- 176 -

(i) SEQUENCE CHARACTERISTICS:

TTAGATATCG GGGCCCACCC TAGCGGT

(147) INFORMATION FOR SEQ ID NO:146:

(i) SEQUENCE CHARACTERISTICS:

(ii) MOLECULE TYPE: DNA (genomic)

(A) LENGTH: 29 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear

PATENT

33

33

AREN-0054

The series

Hart Hark

THE HOLD

٠...

30

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

GGTACCCCCA CAGCCATTTC ATCAGGATC

33